Essays on Early Childhood Development

Pamela Jervis Ortiz

A thesis submitted for the Degree of Doctor of Philosophy (P.hD) to the University College London

September 2016

Declaration

I, Pamela Jervis Ortiz confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Recent research demonstrates that the effects of early childhood environments last a lifetime. There is a vast literature on how parental characteristics and household environment affect investment in children's human capital, but little about how parents' investment decisions and the structure of family dynamics behave. The pathways linking parental characteristics to long-term child outcomes remain unclear. A better understanding of these relationships requires novel modes of inquiry that transcend those of any particular discipline. In this thesis dissertation, I study early skill formation and which factors motivate parental human capital investments by using dynamic behavioural models. Over the four chapters of this dissertation, I address some crucial and unknown research questions as What are the processes (biological, neurological, psychological) that govern the components of human flourishing? How do acquired skills generate new skills and how do they vary at different stages of early ages? What are the determinants of parental investments in children and what are the constraints they face? What are the channels, if any, through which parents' decisions affect child outcomes? Do their decisions respond to incentives/stimulation? Can parents' decisions/behaviour be affected through public policies and by doing so change child outcomes? Doing this, I aim to expand the scope of research on child development to explicitly account for the dynamic interpersonal relationships of attachment, interaction, and scaffolding emphasised in the literature on early child development as well as the fact that it is indispensable to develop more complex economic analysis where preferences, technologies, parental decisions and the importance of dynamics are simultaneously considered in a model. Using models as the ones that I develop in this thesis dissertation it is possible to understand the mechanisms behind decision-making and use them to simulate policies ex-ante that are crucial to addressing all these questions.

Acknowledgments

I would like to thank my supervisor, Professor Orazio Attanasio, for his guidance and support throughout the work of this thesis. Thanks for always being available, for teaching me and for your confidence in me. A special thanks to Italo Lopéz and Flávio Cunha, my co-authors in chapter 3 and 4 respectively of this thesis. I am also grateful to colleagues, professors and staff from the Department of Economics at UCL. I am deeply thankful to Nicolás and my family in Chile for their unconditional support and love.

Funding acknowledgments

This thesis has been made possible through the scholarship "Programa de Formación de Capital Humano Avanzado Becas Chile" from The National Commission for Science and Technology CONICYT, Government of Chile, the Institute Fiscal Studies Scholarship from ESRC, the WM Gorman Scholarship from the Department of Economics at UCL and the "Early Child Development Programs: Effective Interventions for Human Development" from the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Early Childhood Development has been in the centre of the debate in the literature over the last years gave its implications for welfare society. An increasing body of studies in neuroscience, psychology and economics, shows that first years of life are critical for future development of children (Thompson, (2001); Van der Gaag, (2005); Noble et al. (2007); Crawford et al. (2010)). In particular, Heckman et al. (2006) find that a low dimensional vector of cognitive and non-cognitive abilities explain a variety of labour market and behavioural outcomes. Hence, stimulation of these abilities plays a crucial role in the child's future and the social development of future generations. Numerous studies establish that measured cognitive ability is a strong predictor of economic success in life (summarised in Cawley et al., 2001) and those non-cognitive abilities are likely to be an essential determinant of social success in life (Bowles and Ginties, 1976) and for predicting wages, schooling, and participation in risky behaviours (Heckman and Rubinstein, 2001). As a result, there is a significant literature on how parental characteristics and household environment affect investment in children's human capital, but there is little information about how parents' investment decisions, behave: What are the channels, if any, through which these decisions affect child outcomes? Do these decisions, respond to incentives/stimulation? Do public policies influence child outcomes via these decisions?

This is why my research has been focused on the study of those parental investment decisions on early child development, which is crucial information for both, parents and policy makers. Understanding these effects in complex economic environments requires assessing the mechanisms through which they act using simulated counterfactual worlds, and most of the time, incorporating a dynamic point of view to determine long-term effects.

Reducing gaps in multiple dimensions at early life-cycle stages with policies that also promote productivity contributes to shrink social inequality, poverty and exclusion, and to foster economic development. To achieve this goal, the analysis should be twofold: i) identify the factors behind the surge of these gaps, and ii) provide feedback to policy-makers so better policies are implemented to narrow these gaps.

There is wide consensus about the need in social sciences of using rigorous methodologies to assess the impacts of social policies. Recent studies have moved away from more traditional empirical analysis based largely on associations in behavioural data towards the identification of causation and counterfactuals using Randomised Controlled Trials (RCTs). Randomised evaluations can be very powerful tools to obtaining convincing estimates of the average effect of a program or project. There are some examples of well-targeted and well-designed interventions having long-term effects on outcomes (Jamaican home-visiting program (Grantham-McGregor et al. 1991), Perry Preschool Program (Schweinhart et al. 2005) and Abecedarian (Campbell et al. 2002) among others). However, the sole use of RCTs has been criticised on the grounds of external validity and narrow focus of what works, a black-box approach that prevents further understanding why a program or project works and incapacity for examining counterfactuals (Heckman and Smith, 1995; Deaton, 2010).

Structural economic models of individual behaviour, when properly identified, can lead to estimated policy-invariant parameters that govern preferences or technologies, and can be used to learn about behaviours and therefore to evaluate the effectiveness of a program even when that policy has not been implemented. Moreover, structural models provide the tools to understand the mechanisms behind observed decisions in a way that is consistent with accepted theories of economic behaviour. They also provide the flexibility to accommodate observed and unobserved heterogeneity explaining behaviours.

Recently, the most advanced research in the field has turned to the credible estimation of structural models that can inform social policies in more comprehensive ways about relevance, heterogeneous impacts, and the optimal timing of programs.

Human capital investments are affected by both service provision (e.g., health and educational institutions) and by individual and familial decisions, with interactions between them that may mean that the most effective policies may involve improving both service provision and familial inputs. A life-cycle perspective is important because there are likely to be important complementarities over the life cycle, with investments earlier in the life-cycle increasing returns to investments later in the life cycle. International comparisons are informative about how successfully Chile is approaching global frontiers in human capital related to health and skills. Such comparisons suggest, for example, that in many respects such as PISA and PPVT (receptive vocabulary) tests Chile on average performs well in comparison with other LAC countries, but continues to lag considerably behind the European and East Asian countries on the global frontier. Within Chile a high degree of inequality persists in most dimensions of human resources, as well as in income and wealth, with considerable human resource gaps across income groups that start very early in the life-cycle (e.g., 0.8 SD in PPVT scores for 3-5 year olds between top and bottom quartiles) and tend to persist or even increase over the life cycle.

This thesis dissertation aims to provide enough evidence to increase the entire distribution of Chilean human capital towards the global frontier, but at the same time to focus mainly on improving the human capital of the disadvantages Chileans to reduce social exclusion, inequality and poverty. International experiences with various programs over the life cycle are informative about what appears to be "best practice" and therefore merits consideration for the Chile. But best practices from other contexts cannot just be efficiently transferred without adjustments and modifications to Chile because the Chilean context differs from other contexts.

Using models as the ones that I develop in this thesis dissertation it is possible to understand the mechanisms behind decision-making and to simulate policies ex-ante that are crucial to address a set of research questions as: What are the processes (biological, neurological, psychological) that govern the components of human capital? How do acquired skills generate new skills and how does this process differ by the developmental stage of the child? What are the determinants of parental investments in children and what are the constraints they face? Is the lack of knowledge (or awareness) of the potencial returns, the lack for time or monetary resources, or their beliefs the reason why parents behave in one or another way? And what is the relative importance of these constraints? Do public investments in human capital (for example, preschooling or child care services) compensate for or replace parental investments (money and time)?

To address all these questions, it is indispensable to develop more complex economic analysis to simulate the effects of alternative policies and to understand the mechanisms behind decision-making. To do so in the Chilean context, this thesis dissertation provides a better insight of the state-of-the-art of numerical methods and computer technology to develop new techniques posed by modern economic models.

Some of these ex-ante policies to be simulate will be based on current social policies in Chile and associated with early childhood development as extensions of maternity leave, extending parental leave to fathers, parenting programs offered by *Chile Grows with You* as well as a cash transfer (*Asignación Familiar*) to the poorest families.

I divide this thesis dissertation into five chapters, each related to my research questions. In chapter 2, I present a theoretical framework and empirical analysis to contribute to the debate about the determinants of early childhood development in a developing country from Latin America: Chile. The Ecological Environment theoretical model for childhood was proposed to define the determinants of early childhood. This chapter aims to disentangle the determinants behind early childhood development based on multiple empirical strategies through the use of the first and second wave of a recent longitudinal survey, which was designed to characterise the child development. The data contains information about demographics, family's background, cognitive, socioemotional and physical measures for mothers and children under five years old and home assessment environment. The determinants of early childhood development, particularly, cognitive and non-cognitive skills, are studied through the estimation of contemporaneous and value-added cognitive and non-cognitive production functions, as well as the use of factor analysis such as item response theory for reducing the number of inputs. Three main results arise: (1) there are significant socioeconomic gradients in all cognitive tests between poorest and richest quintiles, which lead to a liability among disadvantaged children. Once controlling by observables, the gradient starts to decrease and in some cases to lose significance; (2) there is a significant effect of mother's characteristics and family background at later stage development (above 24/30 months old) measured principally by mother's education, age and cognitive skills, if the family is a two-parent family, the presence of younger/older children as well as home environment measures by parent-child activities, learning materials, parental involvement and verbal and emotional responsibility scores. The later stage development also adds a significant effect on attending a preschool. The previous determinants drive the fall in the socioeconomic gradient in both stages; and (3) regarding the non-cognitive skills, for both waves, the results are similar, there are socioeconomic gradients that are still significant after controlling for all the variables. If the child is male, have a negative and significant effect as well if they attend to preschool. Mother's education and age have positive and significant impact meanwhile having younger children in the household have an adverse and significant effect. Having both parents have a positive impact as well as child's weight at birth and the mother's cognition level. For the first time, all the sub scales of the mother's socioemotional test are (positively) correlated with the child's socioemotional skills. The home environment continues presenting positive and significant effect on child's development.

Chapter 3 is based in a co-authored paper with Italo Lopéz (RAND). We characterise the process of human capital accumulation in early years. Genetics, environment and parental investment at different stages of early years of childhood affect the formation of human capital skills. Only when these channels are adequately incorporated in the study of the human capital formation will be possible to tackle early gaps in childhood and formulating efficient public policies. Despite these recent advances, there is still very little known about the return to cognitive and non-cognitive skills in developing countries. Recent studies have demonstrated how multiple factors relate in a complex way (Cuhna et. al. (2007, 2010)) through the use of technologies of skill formation. We follow the methodology proposed by Cuhna et al. (2010) to estimate a multistage technology of skill formation for capturing different development phases in the early years of a child and dealing at the same time the problem of endogeneity of inputs (correlation with the unobserved shock) and the multiplicity of inputs relative to measures. One contribution to the literature is that we include multiple parental investments not only regarding material resources (monetary investment) and quality time investment but also regarding cognitive stimulation and emotional support. This model provides two critical parameters, the self-productivity of skills (if the child learn how to count, then he can use it to learn other skills which means that skills are self-reinforcing and persist into future periods) and dynamic complementarity (synergy of investments at different t), hence, a second contribution is to analyse if complementarities change with age stages. This chapter also contributes from previous research as include a rich Chilean data to apply the state-of-the-art methodology in the estimation of the production function. Exploiting the rich panel structure of the *Encuesta Longitudinal de Primera Infancia* (Early Childhood Longitudinal Survey (ELPI)) survey we find evidence about the importance of the stock of child's skills as well as early investment in childhood development. Comparing the formation of cognitive and non-cognitive skills in children dealing or not with endogeneity, there is substantial evidence of the effect of parental investment in early childhood development and also support the fact that parental investment is endogenous. Based on the estimation of the same production function but for different age stages, the principal result is how parental investment foster cognitive skills between 24-47 months concerning early and older stages instead for future non-cognitive skills the parental investment have the same effect for all the age stages. There is evidence of cross-productivity for both skills which raises for older stages. Regarding the impact of separating the investment in material resources and quality time in child skills at age t the results show that material resources are essential for determining future child's cognitive skills and quality time for deciding future child's non-cognitive skills. Finally, splitting the investment in cognitive stimulation and emotional support in child skills at age t the results show that there is not much return regarding cognitive stimulation meanwhile the return of emotional support is higher on future child's non-cognitive skills.

formation. We follow the methodology proposed by Cuhna et al. (2010) to estimate a multistage technology of skill formation for capturing different development phases in the early years of a child and dealing at the same time the problem of endogeneity of inputs (correlation with the unobserved shock) and the multiplicity of inputs relative to measures. One contribution to the literature is that we include multiple parental investments not only regarding material resources (monetary investment) and quality time investment but also regarding cognitive stimulation and emotional support. This model provides two critical parameters, the self-productivity of skills (if the child learn how to count, then he can use it to learn other skills which means that skills are self-reinforcing and persist into future periods) and dynamic complementarity (synergy of investments at different t), hence, a second contribution is to analyse if complementarities change with age stages. This chapter also contributes from previous research as include a rich Chilean data to apply the state-of-the-art methodology in the estimation of the production function. Exploiting the rich panel structure of the *Encuesta Longitudinal de Primera Infancia* (Early Childhood Longitudinal Survey (ELPI)) survey we find evidence about the importance of the stock of child's skills as well as early investment in childhood development. Comparing the formation of cognitive and non-cognitive skills in children dealing or not with endogeneity, there is substantial evidence of the effect of parental investment in early childhood development and also support the fact that parental investment is endogenous. Based on the estimation of the same production function but for different age stages, the principal result is how parental investment foster cognitive skills between 24-47 months concerning early and older stages instead for future non-cognitive skills the parental investment have the same effect for all the age stages. There is evidence of cross-productivity for both skills which raises for older stages. Regarding the impact of separating the investment in material resources and quality time in child skills at age t the results show that material resources are essential for determining future child's cognitive skills and quality time for deciding future child's non-cognitive skills. Finally, splitting the investment in cognitive stimulation and emotional support in child skills at age t the results show that there is not much return regarding cognitive stimulation meanwhile the return of emotional support is higher on future child's non-cognitive skills.

Chapter 4 is based on a co-authored paper with Orazio Attanasio (UCL and IFS) and Flávio Cunha (Rice University). We shed light on the importance of maternal subjective beliefs in explaining the heterogeneity in maternal choices of investments in the development of their children. Subjective beliefs about the production function of skills in early childhood development is crucial since parents may have biased beliefs about the returns to investments, which is crucial to pin down in designing policies aimed at remediating poor investments. To determinate the importance of maternal subjective beliefs, we first show how to convert the answers to a specific set of questions into estimates of expected rates of returns on specific investment and then relate these estimates to actual maternal behaviour, then we formulate and estimate a model in which mothers have subjective beliefs about the technology governing the formulation of skills in early childhood development, drawing on detailed and unique data for the identification of the model from an early childhood intervention ran in Colombia, in which, home visitors paid weekly visits to randomly chosen households with the aim of promoting child cognitive and non-cognitive development and improving mother-child interactions. The intervention targeted poor households with children aged 12 to 24 months at baseline and lasted 18 months. We find that parents think that the productivity of investment is much higher for low initial conditions than higher initial conditions. Some findings are worth being discussed. We have elicited maternal beliefs about the production function. We have shown how to relate answers about developmental age under different scenarios to beliefs about returns to investment and parameters of the production function. We find that parents think that the productivity of investment is much higher for low initial conditions than higher initial conditions. We want to extend this approach and estimate simultaneously the production function, the perceived production function and the investment strategy.

In the last chapter, I develop a dynamic structural model estimated with rich longitudinal data from Chile, in which I integrate a children's human capital model with multiple stages of childhood into a dynamic framework to explain parental investment decisions, modeling quality parental investment time and children's technology skill formation accounting for unobserved heterogeneity (income shocks). Parents maximise a constrained model, choosing consumption and quality time with their child and monetary investments in a sequential decision problem using a unitary model.This way, I explore potential mechanisms: First, the effect of parental preferences when they make decisions in each period of a child's life in terms of his/her developmental outcome measure as cognitive and non-cognitive skills; Second, I analyse the constraints parents face when they are taking their decisions in terms of monetary and quality time with their child; and third, the importance of addressing expectations driving investment choices. An important contribution to the literature of child development is a two-step procedure used to eliminate the presence of measurement error in the data for the inputs in the production function as well as integrating a life cycle model into the analysis and hence accounting for the endogeneity (correlation with the unobserved shocks) of investments. In the first stage, I estimate a measurement model based on a linear dynamic factor model and exploit cross-equation restrictions (covariance restrictions) proving that I can identify all of them. In the second step, I estimate together with the dynamic and stochastic structural model that incorporate parental choices based on the overall description of the mechanisms through which parental investment is modified and affects the human capital formation of their children, adding restrictions that involve weaker assumption than those derived from the literature, as well as allowing for simulations of the most effective targeting policies for Early Child Development compensating the most disadvantaged children.

Chapter 2

Disentangling the Determinants of Early Childhood Cognitive Development in Chile

2.1 Introduction

Early childhood development is essential for improving the wellbeing and welfare of society as a whole. An increasing body of studies in neuroscience, psychology and economics, shows that the first three years of life are critical for the future development of children. In particular, Heckman et al. (2006) find that cognitive and non-cognitive abilities are essential to the future performance on a series of social and labour market outcomes: enrolment rates, wages, work experience, crime rates, early pregnancies, drug use in the labour market, among others. Hence, early stimulation of these abilities plays a crucial role in the child's future and the social development of future generations. Heckman and Masterov (2007) show in a cost-benefit analysis that a right investment in early life has a higher economic return than the same investment for adults. This effect is even stronger for most disadvantaged children.

Disentangling the determinants of early childhood development is crucial for understanding how early investment improve abilities in early childhood. Recent studies have shown that focusing on socioeconomic gradient only is far too narrow and hence analysing beyond the existence of such gaps is critical, but not much has been done to address some overarching research questions: what is the causal nature of the early childhood determinants? And what are the key pathways between the early childhood dimensions and cognitive and non-cognitive skills? Accordingly, the objective of this chapter is to explore new theoretical and empirical evidence regarding childhood development and its determinants in Chile.

This chapter, propose the use of an Ecological Environment theoretical model for childhood for disentangling the determinants behind early childhood development. I consider four systems that are related directly to children's development. I use multiple empirical strategies to understand the determinants that predict cognitive and non-cognitive development for children by age stages using children aged between 6-80 months old. The determinants of early childhood development, particularly, cognitive and non-cognitive skills, are studied through the estimation of contemporaneous and value-added cognitive and non-cognitive production functions, as well as the use of factor analysis such as item response theory for reducing the number of inputs. This chapter is one of the few studies that deal with this challenging concern in Chile, mainly due to the inadequate information about the topic. I take advantage of a longitudinal survey designed to characterise children development, using its first two waves, which were collected in 2010 and 2012, and that contains a set of cognitive, socioemotional and anthropometrics measures for both the children and the mothers or primary caregivers. There is also information about home assessment using the Home Observation for Measurement of the Environment (HOME) inventory and the Family Care Indicators (FCI) that help to measure intermediate outcomes and mediators for early childhood development. This information deals with the typical unobserved parental abilities issue, which allows a proper monitoring of early childhood.

Chile is one of the major industrialised countries in Latin America, but at the same time, is the country with greater inequalities among its population regarding income relative to other OECD countries (OECD, 2011). This fact leads perhaps to strong socioeconomic gradients by cognitive and non-cognitive development. Understanding then the nature of the relationship enables formulation and targeting of policy instruments. My contribution is to investigate the early childhood determinants accounting for heterogeneity in the child's environment as well as initial conditions allowing different age-stages and different process behind the technology of childhood skill formation in a developing country.

The chapter is organised as follows. Section 2 reviews the literature in early childhood development for developed and developing countries. Section 3 presents a simple theoretical framework of childhood development to support the empirical analysis. Section 4 describes the data and the measures that are constructed for the empirical analysis based mostly on factor analysis. Section 5 presents the methodology and the empirical results that are used to disentangle the determinants of early childhood development. Finally, section 6 sets the main conclusions.

2.2 A Literature Review

The objective of this section is to present a brief discussion of the most important studies conducted in the spirit of this document. The focus of this literature review is on the determinants of early childhood development. First, I review the international evidence in studies from the developed world, then, I continue with the research from Latin America, and I finalise the literature review with Chile, which is the country of interest in this chapter.

A growing body of studies in neuroscience, psychology and economics demonstrates that the first three years of life are critical for the future development of children. Specifically, cognitive and non-cognitive development begins early in life and is intimately connected with wealth or socioeconomic status (known as socioeconomic gradients), parent-child interactions patterns, child-care centres provision/quality and the learning environment at home, among others. In this sense, international evidence suggests that early childhood development is essential for improving the success of our future children since early development provides the foundation for mental and physical health, and for academic and labour performance.

There are two main branches of literature: one written by neuroscientists and psychologists and the other by economists. The reading of the first suggests that during early childhood it occurs the main development of the brain, which makes early stimulation particularly beneficial. At this stage, environmental stimulation in children results in the generation of new neural connections that alter brain organisation. Consequently, non-appropriate or lack of stimulation at all in early childhood, not only prevents the growth of neural connections but also makes their number diminish progressively. I revise this evidence in what follows.

One of the most cited theoretical references in cognitive development belongs to the psychologist and philosopher Jean Piaget. His Theory of Cognitive Development exposed in the 30's postulates that children go through specific stages as their relationship between intellect and ability to see matures (Chapman, 1988). He mentions four stages: Sensorimotor, Preoperational, Concrete Operations, and Formal Operations. The first stage occurs between birth and two years of age, when children begin to understand the information they receive through their senses and also to develop their ability to interact with the world.[1] The second stage occurs between two and seven years old. Children learn how to interact with their environment in a more complex manner through the use of words and images.[2] Even so, Piaget's approach is a good tool to characterise the cognitive development at childhood, a new generation of neuroscientist argue that this theory neglected brain functions and that it vastly underestimated infant cognitive development (Catherwood, 1999 and Mark Johnson, 1999).

Nowadays, developmental psychologist efforts have been aided by developmental neuroscientists whose conclusions about brain growth complement the findings of behavioural scientists and reflect the importance of the brain in the cognitive development in children. Thompson (2001), states that by the sixth prenatal month, nearly all of the billions of neurones that populate the mature brain have been created, with new neurones generated at an average rate of more than 250,000 per minute. These neurones produce far more synapses with other neurones forming great potential for the developing brain and finished with stimulating experiences activating specific neural synapses, and therefore dropping those that are not enabled progressively over time. Through

[1]Stage marked by Object Permanency, defined as the ability to understand that these objects do in fact continue to exist.

[2]Stage marked by Egocentrism and Conservation, defined as the ability to understand that quantity does not change if the shape changes.

this principle of *use it or lose it* the architecture of the developing brain becomes adapted to environmental stimulation. A similar principle is used to explain the early development of memory ability. Nelson (1995), Diamond et al. (1994) and Dawson (1994) document the growth of early categorisation and thinking skills and early emotional development. It is also possible to establish that early years of children are critical because the brain has its most significant level of *brain plasticity*, which makes the nervous system to show greater resilience and organic and functional reorganisation. This allows children to adapt to the environment and to generate new neural connections by altering brain organisation through influence received from environmental stimulation (OECD, 2007).

Van der Gaag (2005) also finds that newborns have significantly more neurones than a three-year-old child doubling the number they will have as adults. Also, he claims that the brain consists of neural pathways keen to develop particular skills, which if not properly stimulated, do not reach their full potential and are lost. Moreover, a considerable number of recent studies in neuroscience find that the interactive influences of genes and experience shape the architecture of the developing brain, and the active ingredient is the *serve and return* nature of children's engagement in relationships with their parents. This process directly impacts the infant's future productivity, and hence, the social development of future generations (The National Scientific Council on the Developing Child, 2007 and Centre on the Developing Child, 2010). This evidence is consistent with the existence of critical periods in which the brain is particularly effective against certain types of learning. In the case of language, this time ranges from birth to 3 years old. In the case of logical-mathematical, sensitive periods go from 1 to 4 year old (UNICEF, 2004).

Not only neuroscience and psychology studies emphasise the importance of providing appropriate stimulation during early childhood, but also economic research in recent years has studied this phenomenon as a public policy tool that helps improve a number of social welfare indicators. Heckman et al. (2006) analysed in the United States how the development of cognitive and non-cognitive abilities are critical to the future performance of a series of social and labour market themselves: enrolment rates, wages, work experience, crime rates, early pregnancies, drug use in the labour market, among others.

Therefore, it is crucial to analyse two main points in the economic literature. First, if cognitive and non-cognitive abilities genuinely benefit from interventions during early childhood; and second, if this is the case, what are the relevant determinants of early childhood development.

In the first point, the economic literature has focused to show through a cost-benefit analysis, that a right investment in early life has a higher economic return than the same investment for adults (Carneiro and Heckman, 2003). This effect is even stronger for the most disadvantaged children that live in a poor environment (Heckman and Masterov, 2007). In this sense, it is essential to focus on the benefits of short and medium-term interventions generated during early childhood. Currie (2001) states that the benefits can easily compensate between 40% and 60% of the costs of programs implemented on a large scale so that even low long-term benefits are enough to pay the investment. Furthermore, if we do not intervene early on the most vulnerable children, the cost of investing in them when they are adults is so high that it becomes prohibitive (Heckman, 2006; Cunha and Heckman, 2007 and Behrman et al., 2006). There are also studies from organisations that found that when two children born in different families and socioeconomic environments, like those children living in families in the first and fifth quintile, necessarily face different opportunities that lead to different educational and socioeconomic outcomes (OECD, 2009).

Once we establish that cognitive and non-cognitive development begins early in childhood and that the most affected, without early interventions, are the most vulnerable children, the next step is to focus on the effect of socioeconomic status, parental-child interactions patterns, child-care centres, and home-learning environments, among others, as the main sources that affect children's cognitive and non-cognitive development. The economics literature has focused on understanding these variables principally using Regression Analysis. Now, I review the literature that applies this technique to the study the determinants of child's development. The majority estimate the effects of these variables by linear regressions, without taking into account the potential endogeneity issues that arise depending on the variables used. Few studies have tried to deal with this issue using variables to proxy for unmeasured endowments, sibling differences, fixed effects models, experimental/quasi-experimental approach, instrumental variables or using a structural model. Key references include Currie and Thomas (1995); Baydar and Brooks-Gunn (1991); Duncan, et al. (1998); Blau (1999); Duflo (2000); Shea (2000); Levy and Duncan (2000); Ermisch et al. (2002); Duncan, (2003); Mayer (2007); Dahl and Lochner (2008) and Bernal (2008).

Gagné (2003) uses three waves of the National Longitudinal Survey of Children and Youth (NLSCY) to examine whether parental labour market participation and the use of substitute child-care influence the cognitive development of pre-school children using as the dependent variable the Peabody Picture Vocabulary Test-Revised (PPVT-R). The purpose of this study is to determine whether reductions in parental time spent with children impact upon the intellectual development of young children. The author finds that there is no association with school readiness and the number of hours that the majority of pre-school children spend in child-care centres. However, he did observe that children from higher income families using substitute care exhibit better cognitive outcomes than children from lower income families. He also finds that parental labour market participation has little effect on the school readiness scores of most pre-school children. Bernal and Keane (2007) use the 1996 Welfare Reform as an instrument to identify the impact of employment and child-care on the children of single mothers. Using also NLSCY survey, they find that one additional year of child-care lowers cognitive ability test scores by 0.16 standard deviations. This effect depends on the mother, child and child-care characteristics. Noble et al. (2007) use a survey to 150 healthy, socioeconomically diverse first-graders from New York City that were administered tasks tapping language, visuospatial skills, memory, working memory, cognitive control, and reward processing. They show that socioeconomic status explains over 30% of the variance in language and a smaller but highly significant portion of the variance in most other systems. Evans (2004), McLoyd (1998) and Walker et al. (2007) also show the adverse effect of low socioeconomic status on cognitive and non-cognitive child abilities.

Finally, the effect of parental cognitive ability may explain children's cognitive ability. Crawford et al. (2010) use the British Cohort Study (BCS) that provides detailed information on cognitive ability, socioeconomic position, family background characteristics, health, social skills, and attitudes and behaviours across two generations within the same families. They find that parental cognitive ability is a significant predictor of children's cognitive ability; moreover, it explains one-sixth of the socio-economic gap in those skills, even after controlling for a rich set of demographic, attitudinal and behavioural factors.

Most of our knowledge of child development research is conducted in the developed world, in fact, the evidence from Latin America is more limited. This is mostly explained by the availability of high-quality and detailed data on children development. Nevertheless, there is important and substantial evidence found.

In Mexico, Gertler and Fernald (2004) analyse the impact of the *Oportunidades* program designed to target poverty by providing cash payments to families in exchange for regular school attendance, health clinic visits, and nutritional support. They find that poorer children in their sample appeared to have severe cognitive deficits in long and short-term memory, vocabulary, and visual integration compared to the population used to norm the Woodcock-Johnston-Muoz and the PPVT.

In Ecuador, Paxton and Schady (2005) examine the relationship between early cognitive development, socio-economic status, child health, and parenting quality (using Home Observation for Measurement of the Environment (HOME) scale). They use a sample of over 3,000 predominantly poor preschool age children and analyse determinants of their scores on PPVT Spanish version (TVIP). They estimate that a child whose family is in the ninetieth percentile for wealth, maternal and paternal education is associated with a language score that is approximately two standard deviations in test scores higher than a child whose family is in the tenth percentile. They also show that parenting quality is associated with child cognitive development. Children who live in households with fewer siblings have higher test scores, as do children whose parents receive better scores on an index of the quality of the home environment.

In Argentina, the Centre for Research on Child Nutrition (CESNI) and the Córdoba Lactation, Feeding, Growth and Development Foundation (CLACYD) conducted a study by 518 children aged 5 to 28 months old, 5 years old and 8 years old. The results showed that for the 5 years old group the malnutrition chronic reached 2.9%, with a higher percentage in the low category of socioeconomic status. The environmental stimulation variable proved to be the most influential in cognitive performance with a difference of 48 points in the scales of the HOME between the low and high socioeconomic status (CLACYD, 2000 and CLACYD, 2002). In another study in the same country, Piacente et al. (1990) use a survey of 1,521 children less than 5 years old and 920 mothers with characteristics of families in marginal areas to determine the psychological development and nutritional status of children. The results showed, when they compare the performance using a control sample consisting of non-poor children, that they had higher scores in the different areas of psychological development, particularly in language.

In Paraguay, Peairson et al. (2008) study the effect of the *Pastoral del Nio* program, which served a non-random sample of children aged 24 months old or younger from poor rural Paraguayan areas. They show that program children scored significantly higher than non-program children on the mental development index portion of the Bailey Scales of Infant Development II test at 0-4 months old and also at 20-24 months old. The most important variables that explain this result include health, nutrition, and education variables.

In Brazil, Halpern et al. (1996) use a sample of 20% (1,400 children) of all children born in 1993 in hospitals from the Rio Grande do Sul, Brazil, these children were followed through home visits during the first year having nutritional status and Denver II Test. A 34% of the children assessed at 12 months failed this test, where failure was associated with birth weight and family income variables that were strongly related to the potential risk of developmental delays at the age of 12 months old.

As reviewed in this section, the major problem in achieving an excellent cognitive and non-cognitive development is the existence of socioeconomic gradients that affect all the others variables that determine the socioeconomic status. They included parental-child interactions patterns, physical status, child-care centres, and home-learning environment among others.

Today, Chile is one of the major industrialised countries in Latin America. In fact, in 2007, it achieved the status as the region's richest country regarding GDP per capita. It also has low unemployment rates, a stable monetary policy, political transparency and stability. However, according to the OECD (2011)

Chile is the country with greater inequalities among its population regarding income relative to other OECD countries.[3]

In this context, public policies aimed at early childhood are intended to tackle the problem of inequality. Since 2006, the country experienced two policies that are designed to improve protection of children in Chile that is essential to the future of vulnerable children. The international community, through UNESCO, highlights this fact in its "Education for All Global Monitoring Report" published in 2010 where it is claimed that Chile has begun to implement a strategy for the development of children focused on health and education. Its purpose is to provide early childhood care and education to all children under five years, focusing primarily on those belonging to the two poorest quintiles.

The first element is the increase in the supply of childcare centres nationwide, while the second is the implementation of *Chile Grows with You* program (Vega, 2011) that provides free access to childcare centres and parenting advice for children among the poorest 60% and between3-months-old to 4-years-old. A study by the United Nations Development Program (UNDP) examines the implementation of the program and its effect on the female labour market. It can be concluded that the effect of increased access to child-care centres on female labour supply is essential and results in considerable improvements in household income and levels of poverty (UNDP, 2008).

Another line of work that looks at improving the knowledge of childhood development is to increase the availability of high quality and detailed data on children. A partnership formed in 2007 between the *Junta Nacional de Jardines Infantiles* (JUNJI), an organisation of the Ministry of Education and the *Centro de Estudios de Desarrollo y Estimulación Psicosocial* (CEDEP), set out to implement a longitudinal study to evaluate the effects of participation in the child care program on child's development outcomes as measured by the Spanish version of the Battelle Developmental Inventory. Characteristics of a child's family and private child-care establishments were gathered at the beginning of the study. Another study from JUNJI, UNICEF and UNESCO

[3]The OECD noted that the Gini coefficient that measures inequality in Chile of 0.50 when the average organisation is 0.31. Also, 18.9% of Chileans are poor; a number surpassed only by Mexico (third) and Israel, far from the 10% across the OECD.

is the *Encuesta Nacional Primera Infancia* survey which goal was to collect information about early childhood (from birth to 5 years 11 months old) for describing the development of 6,500 children to respond relevant public policies. The most important source of data on children is the *Encuesta Longitudinal de Primera Infancia* (Early Childhood Longitudinal Survey (ELPI)) survey, which is a longitudinal study and allows researchers to assess the impact of early childhood policies and provide valuable information for the evaluation and design of social policies in this field. It is based on a representative sample of 15,000 children under 5 years old and their families for the first wave, during the second wave, the sample is 18,000 as included a refreshing sample from 0 to 3 years old. This survey provides demographic information, measures of cognitive and non-cognitive abilities for children and their mothers, anthropometrics and the quality of care for children provided at home among others.

Noboa and Urzúa (2010) use quasi-experimental methodologies and took information from the Longitudinal Survey conducted by JUNJI in 2007. They determined through a series of indexes of cognitive and non-cognitive abilities, the average effect that has on children who attend to public child-care centres. They use a random sample of 41 public child-care centres, and the treatment group was randomly selected from children. The final sample consisted of 331 children between 5 to 14 months old, attending a public child-care centre in April 2007 and the control group was randomly selected from healthy children who did not attend any child-care centre but attended nearby health clinic finished with a sample of 151 children. The purpose was to match socioeconomic conditions across groups (roughly the same age). They found that while attending the initial effect was negative, over time the effect on all areas of development measures was positive.

Another study with considerable less quality in data is one written by Riquelme del Solar (2003). The author created a test on basic abilities for calculus introduction (TIC) for children between 5 to 6 years old. The study obtained that 32 items according to five cognitive abilities constitute this instrument proposed by J. Piaget. The author uses a sample of 60 children attending public and private preschool in equal proportion. Even with the observed data, the study was able to display immediately that a child belonging to upper middle socioeconomic status is less strong in the notion of conservation abilities. In contrast, children belonging to low socioeconomic status are strengthened deprived in classification and seriation ability. They focus their weaknesses on conservation, expression and symbolic logic trial function.

Cárcade et al. (2015) use two studies, the Magellan-Leiden Childcare Study (MLCS) and the ELPI to study if Mapuche and non-Mapuche parents provide substantially different child-rearing environments for their children. They find that the differences between the two ethnic groups are explained by income, which confirms the important role that socio-economic factors play in child rearing and parenting. It is concluded that preserving cultural differences and traditions is an important goal in itself but that for childcare in Chile it seems equally important to eliminate socio-economic inequality.

Coddington et al. (2014) use only the first wave of the ELPI to study the relation between family socioeconomic status and children's receptive Spanish vocabulary, they find evidence of partial mediation through indices of the standard of living and parents' level of cognitive and linguistic stimulation in the home. Contreras et al. (2015) estimate the effect of variables related to health status, cognitive abilities, and demographic factors of Chilean mothers and children on the children's psychomotor development and they find that health, cognitive, and demographic variables are important factors in a child's biopsychosocial development.

This paper is among the first to estimate early childhood production functions by age stages and by multiple skills in a context of a developing country.

2.3 A Simple Theoretical Framework of Childhood Development

Throughout this section, I present a theoretical framework that places the variety and interrelation of all factors that surrounds the evolution of cognitive and non-cognitive abilities in childhood. It takes into account that this is an essential tool for understanding how is possible to promote successful lives for children and therefore give some guidance for focusing public policies on early intervention efficiently and accurately.

I explore some of the theoretical models that met the basic premise of covering as far as possible, as many variables that can affect in some way childhood development. The review led to the model by Bronfenbrenner's ecological (1979), which offers one of the most comprehensive and effective ways to address this issue. The author advocates a systemic, naturalistic and integral vision for the understanding of physiological development as a complicated process that responds to the influence of multiple factors that are closely linked to the environment or surroundings (*ecological environment*), in which this development takes place. The predecessor in this thought, Lewin (1936), stated that the conduct of the exchange is a function of the person with the environment.[4] However, Bronfenbrenner adds the idea of human development as a progressive accommodation between being active human, who is under development, and their immediate surroundings. This process is, also, influenced by the relationships between these environments and the broader contexts in which those environments are included. I describe each of these systems and I focus on the proposed implementation of themselves to childhood development from the postulates of the author. Specifically, there are four levels or systems that directly and indirectly affect the children's development.

The first system is called Microsystems, and it corresponds to the pattern of activities, roles and interpersonal relationships. They develop the person's experiences in a determined place in which it participates (the closest level to the subject). It includes behaviours, roles and relationships that are common daily contexts in which he/she spends his/her days. In the case of childhood development, this system can be called *Child System*. There is international agreement to conceive that this process occurs at least in 5 areas: motor, physical growth and maturation, cognitive, socioemotional, and social areas. In fact, from birth to five years old, children develop skills on which development is built it. This can be seen in motor, cognitive and language abilities and their emotional, social, regulatory and moral progress. These critical dimensions of early development are intrinsically linked, and each requires focused attention. As reviewed in the literature, every aspect of early development from brain

[4]Which he expressed in the equation C (behavior) = Function (P(person)A(environment)).

circuitry to the capacity of children's empathy is affected by the environment and the accumulated experiences, that start early in the prenatal period and that extend through the years of early childhood (Nelson, 1995; Diamond et al., 1994 and Centre on the Developing Child, 2010 among others).

The second system is called Mesosystems, and it includes the relationships between two or more settings in which the developing person participates. It contains the interrelations of two or more environments where the developing person actively participates. Therefore, it is a system of microsystems. These systems form when the person enters into a new environment. For childhood development, this system could be called *Family System*. Parents and other regular caregivers are essential assets of the influence of the environment during early childhood. Children grow and thrive in the context of relationships that provide love, security, responsive interaction, and stimulation for exploration. Without these relationships, the development is disrupted, and the consequences can be severe and permanent. Hence the early development of children depends on the welfare of their parents. With regard to the influence of parenting, it is essential to consider that several legacies are given to children and daily routines for such ordinary activities as sleeping, eating, playing with the values that prevail in their conceptions of discipline, role gender, religious or spiritual values among others, and the contexts that frame the cognitive, linguistic, and socioemotional and therefore influence the acquisition of specific abilities or behaviours (Shonkoff and Phillips, 2000).

The third system is called Exosystems and refers to one or more environments in which the person in development is not included directly, but to those environments in which events that affect the person. For childhood development, this system could be called *Environment/Neighbourhood System*. It refers to the physical environment, culture, values, social capital, social networks, and geography, among others that are involved in child development. They include health system, child-care centres, preschool and primary establishment, basic services, the workplace of the parents, the circle of friends of parents, among others.

The last system is called Macrosystem and refers to cultural and ideological frameworks that affect or transversely affect the other three systems. It gives certain uniformity in form and content and also some difference from

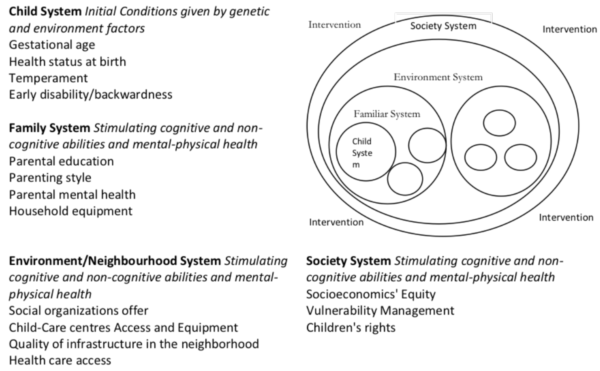


Figure 2.1: Ecological Environment for Childhood Development

other environments influenced by other dissimilar cultural or ideological frameworks. As Bronfenbrenner notice, in society, the structure and substance of the three other systems tend to be similar, but among different social groups, the constituent systems may have notable differences. Therefore, analysing and comparing the child, family and environment/neighbourhood systems that characterise different social classes, ethnic and religious groups, is possible to systematically describe and distinguish the ecological properties of these social contexts. In the case of Childhood Development, this system could be called *Society System*. Latin America and particularly Chile still have two significant issues that still generate differences among children: income inequality and sex discrimination. Relative to the first, Chile is the country with higher inequalities among its population regarding income with respect other countries OECD members. In the second issue, Chile has cultural or ideological frameworks that have effect in the child on the gender archetype that is developed through the biological roles and identity, cognitive and social influence by sex stereotypes. In this context, it is possible to provide tools that detect cases of violence and promote social activities for children under conflict resolution based on dialogue without discrimination. UNESCO (2010) finds that all this gender-based violence is the result of what children experience daily in their homes. They notice that women, mothers and girls do more tasks inside and outside the home than men and children. This is the foundation of the problems detected by scholars.

Figure 2.1 shows the four systems that are the basis for cognitive abilities, language acquisition and empathy with other humans. In other words, they are the pillars of childhood development. Once the child, familiar, environment/neighbourhood and society systems are understood, it is possible to disentangle the determinants of early childhood development and therefore an efficient and correct intervention.

2.4 Data and Measures

One of the major limitations for disentangling the determinants of childhood development in Latin America and particularly in Chile is the lack of high quality and detailed data on children under 5 years old. In this section, I present new data that allows researchers to assess the impact of early childhood policies and to provide valuable information for the evaluation and design of social policies in this field since this data contains variables determining children's cognitive and non-cognitive abilities. Throughout this section, I develop an empirical analysis that allows me to generate relevant measures of early childhood using Principal Components Analysis, Factor Analysis and Item Response Theory.

2.4.1 Survey Description

The Early Childhood Longitudinal Survey is a longitudinal survey that collects information about children and their families in the first few years of the child's life. The first wave was conducted in 2010 and the second one in 2012. It is a representative survey of children from urban and rural areas who were born on January 1, 2006, and August 31, 2009, in the first wave, and that is a representative sample. The sample includes different cohorts of children, distinguished by year of birth. The second wave follows the first wave but also add a refreshing sample of around 3,000 children born on September 1, 2009, and December 31, 2011[5].

The sample design corresponds to a stratified sample, where strata are constructed by clustering of districts that had similar socioeconomic status.

[5]The sampling frame for the first wave corresponds to 1.297.822 birth records from March 1, 2004, and August 31, 2009, however, the final sampling frame used information from January 1, 2006, and August 31, 2009 (total 44 months) due to the test's timing.

The selection of the units of analysis was made using systematic random sampling. The distribution of the sample was made in proportion to the population of each stratum considering the country's 15 geographical regions (Levy and Lemeshow, 1999). A representative sample of children was recognized as the best option, using a stratified two-stage (communities and then children) design for clusters (Kish, 1965).

Module	Description
A	*Household Composition*: It contains information that identifies whether the person that answers the survey is the biological mother or not, and if the selected child has a twin-brother. It also contains the most important socio-economic characteristics of the household's members, which includes the relationship with the selected child, age, gender and marital status, among others.
B	*Education*: It contains information on the education level and the administrative characteristics of the child-care establishments of each household's members.
C	*Employment Status*: It contains information on the occupational status and job characteristics for of each household's members that are more than 15 years old.
D	*Household Income*: It contains information on the income and salary for each family member (which also includes subsidies, pensions and rent, among others).
The following questions are made only for the interviewee (usually the child's mother or caregiver)	
E	*Social Protection*: It contains data on access to the health system.
F	*Assets and Equity*: It contains characteristics of the home as artefacts and/or services, type of housing, predominant construction material and the number of rooms, among others.
G	*Mother's pregnancy*: It contains information (twelve questions) on biological mother's health status during pregnancy (pregnancy control, diseases, medical conditions, nutritional status, certain circumstances that may have occurred during pregnancy, etc.). The nexr questions include information about birth (birth establishment and condition, complications during birth, months that the child was breastfed, etc.).
H	*Learning and Chile Grows with You*: It contains information about activities, games, learning materials used by the child, and about the participation in the Chile Grows with You program.
I	Immunization: It contains information from the child's immunization (vaccination information).
J	*Child Care*: The module is performed for 8 time periods for the child, starting when he/she is 0-3 months old to 5 years old of age. It contains information related to child-care centres and mother's employment.
K	*Selected Child's Biological Father*: It contains information about the biological father in terms of educational, occupational and socioeconomic status.

Figure 2.2: Socio-Economic Questionnaire

The survey contains two major areas in both waves. The first is a household questionnaire divided into two main sections. The first one (Modules A, B, C and D) contains questions designed for each household members, while the second part (modules E, F, G, H, I, J and K) contains questions that apply only to the person that answer the questions, i.e., the mother or the primary caregiver of the target child. Table 2.2 presents the information contained in each module for the first wave, the modules are the same for the second wave, but primary caregiver's labour history was included.

Also, the second area of the survey collects a set of cognitive, socioemotional and anthropometrics measures for both the children and the mothers or primary caregivers. There is also information about home assessment using the Home Observation for Measurement of the Environment (HOME) inventory and the Family Care Indicators (FCI) that help to measure intermediate outcomes and mediators for early childhood development. Tests applied in both waves measured the development of children in different areas, such as motor, cognitive, language, emotional and social areas. Evaluate the overall development of children allows identifying the areas in which they have higher and lower achievements. This information is relevant at the time of reporting social and educational policies.

The following describes the tests applied to study the overall development of children evaluated and the primary caregiver in both waves of the survey.

Cognitive Measures

Escala de Evaluación de Desarrollo Psicomotor (EEDP) Psychomotor skills (for children 0-23 months old)[6], not in English, could be translated as Psychomotor Development Rating Scale. The test measures the performance and the reaction of the child to certain situations to be resolved for which a certain level of psychomotor development is required. A child whose psychomotor development is by what is expected for their age should get a ratio close to the average development (100) to a standard deviation (85) (Rodríguez et al., 2008). Was applied only in the first wave.

Battelle Developmental Inventory (BDI 2010) Psychomotor skills (for children 0-23 months old). The BDI contains 341 items across five develop-

[6]The information in the parenthesis shows the age range for which the test was applied in the survey

mental domains: personal-social, adaptive, motor, communication, and cognition. These five domains are further divided into twenty-two separate subdomains. The personal-social domain is composed of: adult interaction, expressions/feelings/affect, self-concept, peer interaction, coping, and social role. The adaptive domain includes attention, eating, dressing, personal responsibilities, and toileting. The motor domain is composed of muscle control, body coordination, locomotion, fine muscle, and perceptual motor. The communication domain includes: receptive and expressive. Finally, the cognitive domain is composed of: perceptual discrimination, memory, reasoning/academic skills, and conceptual development. The test results can be analysed using z and T scores and the ratios of standard deviations as the basis for conclusions concerning the strengths and weaknesses of development (De la Cruz and González, 1998). Was applied only in the first wave.

Battelle Developmental Inventory, Screening Test, 2nd ed. - (BDI 2012) Psychomotor skills (for children 6-84 months old). The BDI 2012 includes the same areas as the BDI 2010. Has 96 items (two per each age level) extracted from the full version of the BDI and it is a screening test that evaluates the child development from 0-8 years old. The objective is to assess the fundamental skills development in five areas (personal and social, adaptive, motor, communication and cognitive). Was applied only in the second wave.

Test de Vocabulario en Imágenes Peabody, Hispanic America adaptation (TVIP) Language skills (for children 30-84 months), is the Spanish version of the Peabody Picture Vocabulary Test (PPVT). This test measures the child's comprehension and understanding of vocabulary using relating words to an illustration. The scale should be viewed primarily as an achievement test since it shows the extent of Spanish vocabulary acquisition of the subject. Also, it may be considered to be a screening test of scholastic aptitude (verbal ability or verbal intelligence), or as one element in a comprehensive test battery of cognitive processes when Spanish is the language of the home and community into which the subject was born, has grown up, and resides; and when Spanish is, and has been, the primary language of instructions at school (Dunn et al., 1986). Was applied only in both waves.

Test de Desarrollo Psicomotor (TEPSI) Psychomotor skills (for children 24-60 months old), not in English, could be translated as Psychomotor Development Test. TEPSI is a screening test, i.e., an assessment list showing the level of performance in the psychomotor development of children between two and five years old, in terms of a statistical norm established by age group and whether this performance is standard, or is under expectations through the observation of the child's behaviour in situations proposed by the examiner. The TEPSI yields results at global as well as sub-scale levels, which are coordination, language and motor functions (Haeussier and Marchant, 2003 and Wechsler, 1974). Was applied only in the first wave.

Test of Learning and Child Development (TADI) Psychomotor skills (for children 6-80 months old), not in English. TADI is a Chilean instrument that allows measuring what children know, and what they do, according to four dimensions of development: language, cognition, motor and socio emotionality, each of which constitutes a separate scale. Therefore, the TADI allows to evaluate learning and development globally, covering the four dimensions, or by each dimension separately (CEDEP, CIAE, 2012). Was applied only in the second wave.

Wechsler Adults Intelligence Scale (WAIS) Intelligence/cultural level and memory skills/processing speed/short-term auditory memory (for primary caregivers) through the Vocabulary and Digit Span sub-scales respectively. The WAIS measures human intelligence reflected in both verbal (which measures the subject's knowledge of word meaning) and digital distance (ability to recall digits from memory, performance based on the maximum length of a list of digits the subject can remember) abilities. It is based on the belief that intelligence is a global construct, which reflects a variety of measurable skills and that can be considered in the context of the overall personality. It has been demonstrated that the test provides highly accurate measurements and has a high predictive capacity regarding the future behaviour of an individual. The WAIS is also administered as part of a battery test to make inferences about personality and pathology; both through the content of specific answers and patterns of subtest scores (Apfelbeck and Hermosilla, 2000). Was applied only in the first wave.

Social-Emotional Measures

Child Behaviour Checklist (CBCL1) (for children 18-60 months old) Obtain standardised ratings, and descriptive details of children's functioning, as seen by parents/caregivers providing results for three general scales: Total Problems, Internalising and Outsourcing There are 7 syndrome scales designates as Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems, and Aggressive Behaviour (Achenbach and Rescorla, 2000). Was applied only in both waves.

The Big Five Inventory (BFI) (for all primary caregivers) The BFI is a self-report inventory designed to measure the Big Five dimensions. The five factors are Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. It contains 44 items and consists of short phrases with relatively accessible vocabulary (Casullo, 2000). Was applied only in the first wave.

All Cognitive and Social-Emotional measures have a raw score as well as T scores. As the analysis is within the sample I have internally standardised each measure, taking advantage of the sample size per month, using non-parametric estimation for age and hence I remove the age effect. For this, the first step is to use kernel-weighted local polynomial smoothing methods to regress the raw score on child's age (in months) and recover the mean. The second step is to use again the same estimation method to regress the square of the residuals in the first regression on child's age and recover the variance. Finally, for calculating the z-score, I subtract the mean and divide by the standard deviation for each raw score.

Figures 2.3 and 2.4 shows the z-scores for the cognitive and socioemotional tests.

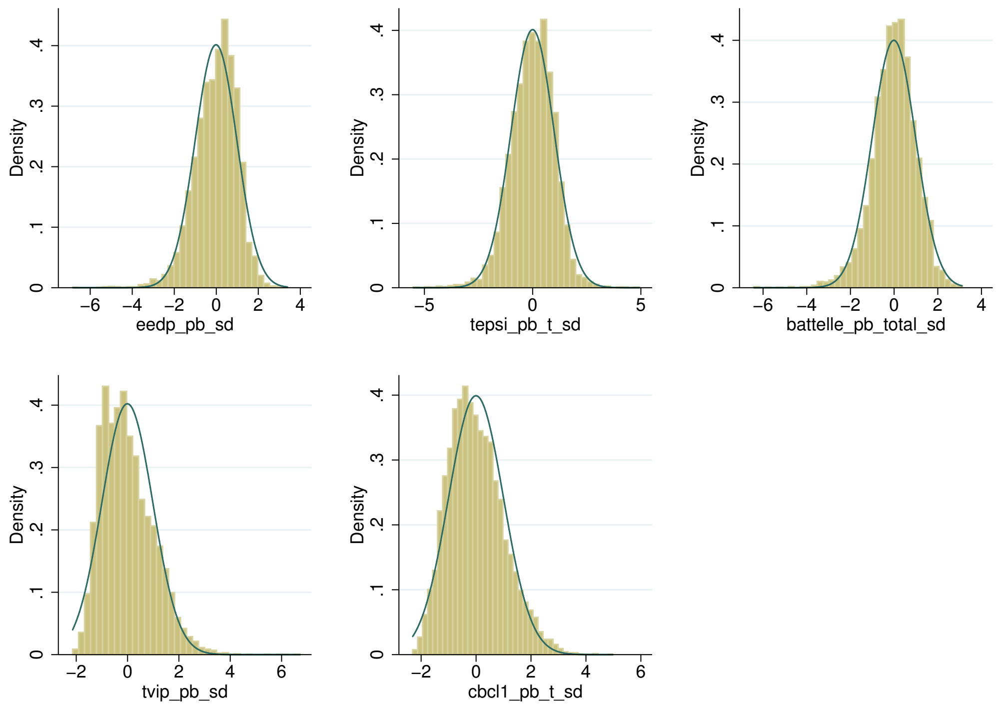


Figure 2.3: Child's Cognitive and socioemotional Measures, ELPI 2010

Anthropometric Measures at birth

Height and weight for children were converted to z-scores using the World Health Organization (WHO, 2006). The results can be classified as delayed, normal or increased, according to its z-score of the curve established by WHO where the normal category is when the measure is between -2 and +2 z-score, increased/delayed with z-score bigger/smaller than 2/-2.

Home Assessments

The Home Observation for Measurement of the Environment (HOME) inventory (Bustos et al., 2001) measures the quality of stimulation and support given to a child in a family atmosphere. The inventory was created by Caldwell et al. (1984, 2003). It consists of 55 items grouped into eight subscales, which records the presence or absence of the trait. This score is obtained from

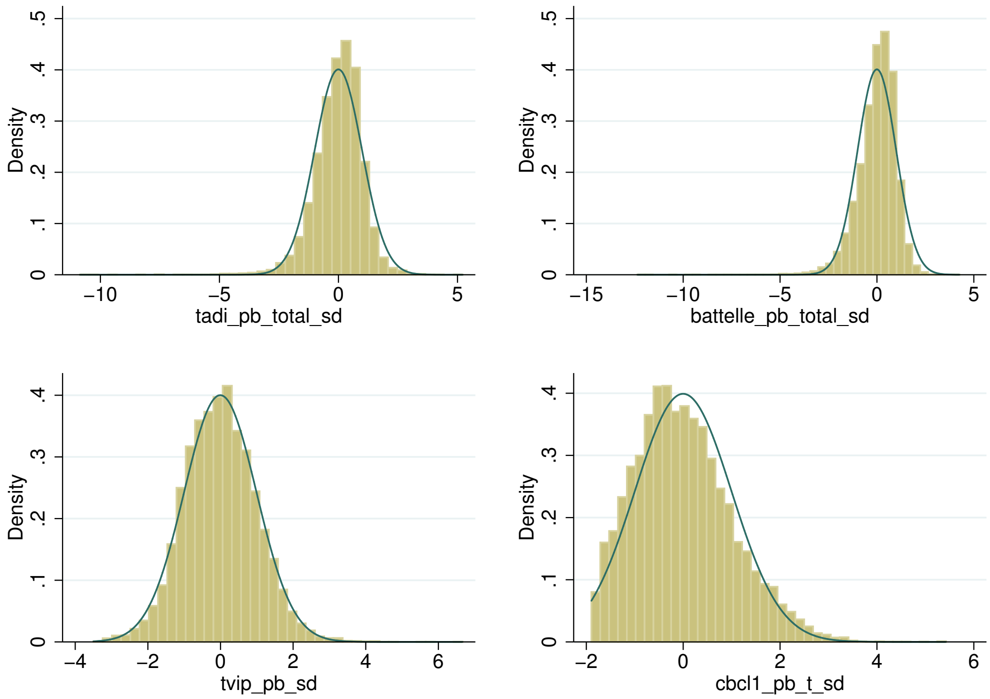


Figure 2.4: Child's Cognitive and socioemotional Measures, ELPI 2012

a combination of observation, and a semi-structured interview is conducted in the child's home with the presence of the mother and child. The inventory has eight subscales: Learning Materials, Stimulation of language, Physical environment, Responses from parent to child, Academic stimulation, Modelling and stimulation of social maturity, Diversity of experiences and Acceptance of the child. Nevertheless, in this survey was only collected some items that could be observable during the interview generating new subscales that I will review in the next subsection. Finally, were added three items to observe eating behaviours.

The Family Care Indicators (FCI) (Hamadani et al., 2010) is a survey-based indicator of the quality of stimulation of the home environment that help to measure intermediate outcomes and mediators for early childhood development. The FCI was developed to measure home stimulation in large populations and were derived from the Home Observations for Measurement of the Environment (HOME). It has items that can be related to i) parent-child interactions as reading books, singing songs, taking the child outside the home, playing with the child, spending time with the child, etc. and ii) learning materials.

2.4.2 Factor Analysis

In this subsection, I develop an empirical analysis to generate relevant measures for analysing early childhood using Principal Components Analysis (Pearson, 1901 and Hair et al., 1987), Factor Analysis (Child, 1973) and Item Response Theory.

Principal Component Analysis (PCA) is defined as a linear combination of variables used in the analysis. In this method, the new variables that are generated are independent of each other and hence are uncorrelated indexes or components. With this technique, it is possible to reduce the number of variables that we have originally. The newly created components can explain much of the total variability of the data. The weights for each principal component are given by the eigenvectors of the correlation matrix, while the variance for each principal component is given by the eigenvalue of the corresponding eigenvector (Hair et al., 1987).[7]

Alternatively, Factor Analysis (FA) is defined as a latent variable model, not as PCA but that aspires to reduce the number of variables. In this case, the factors (instead of the components) derived from the analysis are assumed to represent the original processes that result in the correlations between the variables. The factors are continuous and follow a multivariate normal distribution. In FA, it is necessary to choose an option for estimation among the principal factor (default option), the iterated principal factor (same as the principal factor but that is computed iteratively), the principal component factor (same as PCA), and maximum likelihood. For the two methodologies (PCA and FA), the number of extracted components/factors is defined by a couple of common methods used to select components/factors. The first method is a screen plot looking for the elbow in the chart, while the component/factor

[7]Notice that categorical data are not suitable for PCA, as the categories are converted into a quantitative scale which does not have any meaning and also it is important to notice that if the data has been standardised, then PCA should use the covariance matrix.

before the elbow is the number of components/factors to keep. The second one for deciding is to just keep every component/factor with the associated eigenvalue greater than one. The logic here is that each variable accounts for a variance of 1 so if a component/factor accounts for a variance of more than 1, then it accounts for more variance than any one of the original observed variables could (Hair et al., 1987).

The classical analysis of items does not provide information about how the level of ability interacts with the performance required by the item to be answered correctly. This situation is resolved by the implementation of Item Response Theory (IRT). The Item Response Theory assumes that responses to the items of a test or inventory can be explained from a latent trait, or that the underlying latent trait can be explained from the answers to the items of a test or inventory. One of the key concepts of Item Response Theory is the Item Characteristic Curve (ICC), which is represented by a graph that shows the probability of responding correctly to an item as a function of the performance of the underlying latent trait in an item from the test or inventory. The ICC is a curve which increases a person's ability increases since it has been more likely to answer the item correctly. The importance of the ICC regarding the classic indicators of difficulty and discrimination is that it allows visualising how the probability of responding correctly to an item depends on the level of the latent trait that the evaluated has. The application of Item Response Theory is valid under the fulfilment of the following assumptions:

- Local Independence: The probability of responding correctly to the item interacts only with the skill level, and not the result of another factor (tracks that give other items within the test or inventory).
- Unidimensionality: It is defined regarding statistical dependence between the items of a test or inventory. Specifically, the requirement that a test/inventory is unidimensional is that the statistical dependence between items can be explained by a dominant latent trait. This means that a test is unidimensional if their items are statistically dependent for entire testing population, and for a dominant latent trait.
- The latent trait is continuous and normally distributed.

- Conditional on the latent trait, the responses to any two items are independent of each other.

The most popular models of the Item Response Theory are logistic models, which differ by the number of parameters to be estimated and can only be applied to binary responses items. I use the standard two-parameter logistic (2PL) IRT model for recovering the underlying latent trait which assumes that the interviewee did not use random to answer an item correctly. Therefore, the probability of correctly answering an item depends only on the skill level of this person and the characteristics of the item, which are: difficulty and discrimination, since it is assumed that the random tends to zero. In particular, let the probability of the person i having a value of 1, which indicates a positive or "correct" response, for item j be:

$$Pr(Y_{i,j} = 1 \mid \alpha_j, \beta_j, \theta_i) = \frac{e^{\alpha_j \theta_i + \beta_j}}{1 + e^{\alpha_j \theta_i + \beta_j}}, \tag{2.1}$$

where θ_i denote the underlying latent trait by interviewee i. The parameter α_j represents the discrimination of item j which tells how fast the probability of success on that item responds to small changes in the latent trait, close to the difficulty of that item. The higher the discrimination parameter, the better that item discriminates, around its difficulty value, between interviewees of similar levels of the latent trait. Using the IRT parameterisation; the transformation is $a_j = \alpha_j$ and $b_j = -\beta_j/\alpha_j$, where the parameter b_j is the difficulty. The item is more difficult the higher it's level of b_j and therefore the lower the level of β_j.

Based on a factor analysis, three measures will be studied below in the assumption that these are associated with early childhood development as shown in Section 2.3: Wealth factor, SES factor, and Home Assessments factors. For the first two, the idea is to construct two factors that measure the same (wealth and socioeconomic status factor). The chosen factor has the best performance according to principally to the Kaiser-Meyer-Olkin measure of sampling adequacy.

2.4.2.1 Wealth score

Economists have followed the proposal made by Filmer and Pritchett (2001) to use PCA to aggregate several binary asset ownership variables into a single dimension. Nevertheless, Kolenikov and Angeles (2009) show that the currently used method of running PCA on a set of discrete variables does not perform as well as other methods for analysing discrete data (which include the use of ordinal variables or polychoric correlations). Polychoric PCA uses maximum likelihood to calculate how that continuous variable would have been split up to produce the observed data.

Therefore, I follow the standard approach of using Polychoric PCA[8] to construct a wealth score to combine binary, ordered categorical and continuous data. The score is constructed using information on characteristics of the home as artefacts and/or services as refrigerators, microwave ovens, washing machines, cell phones and internet, among others (binary data), predominant material in housing (ordered categorical data), and the rent that is paid or that would be paid for the bedroom/housing (continuous data). As there are several dependent variables in the empirical analysis, each one with a different number of observations, I construct a wealth score for each sample, where the first principal component explains in average 49% of the variation in these variables (using multiple imputation techniques before generating the score when there are missing values). The wealth score was divided into quintiles where the last quintile is the richest. Appendix 1 shows the relevant statistics and figures for the scree plot criterion using the first wave of the survey.

Figures 2.5 and 2.6 show the z-scores for the cognitive measures by child's age for the poorest and richest quintile.

[8] Independently of the factor strategy all methods create similar scores. The pairwise correlation between PCA, two FA (because the principal component factor is inappropriate due is based on the assumption that the uniquenesses are 0 but there is considerable uniqueness, and the principal factor method has negative eigenvalues) and Polychoric PCA is about 0.99 for each sample. Appendix 1 shows these results.

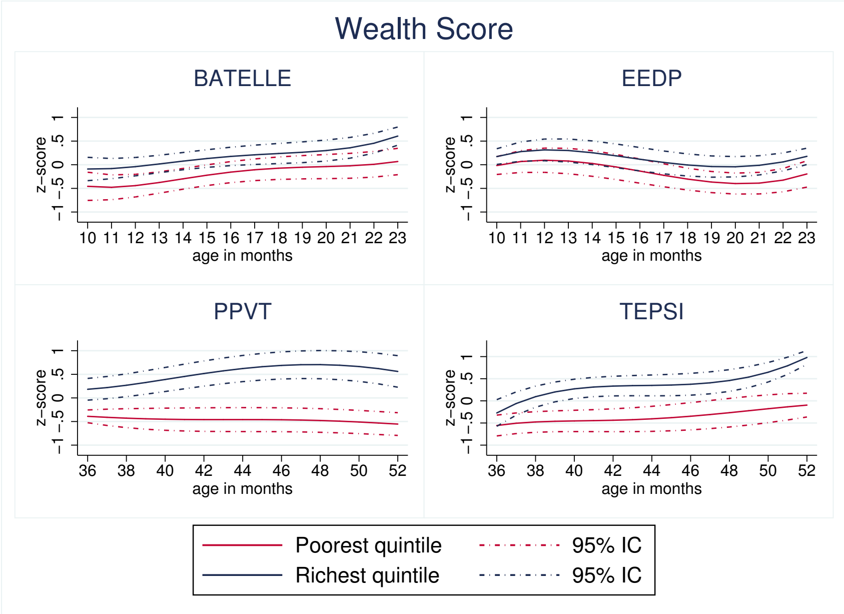


Figure 2.5: Child's Cognitive Measures and Wealth score, ELPI 2010

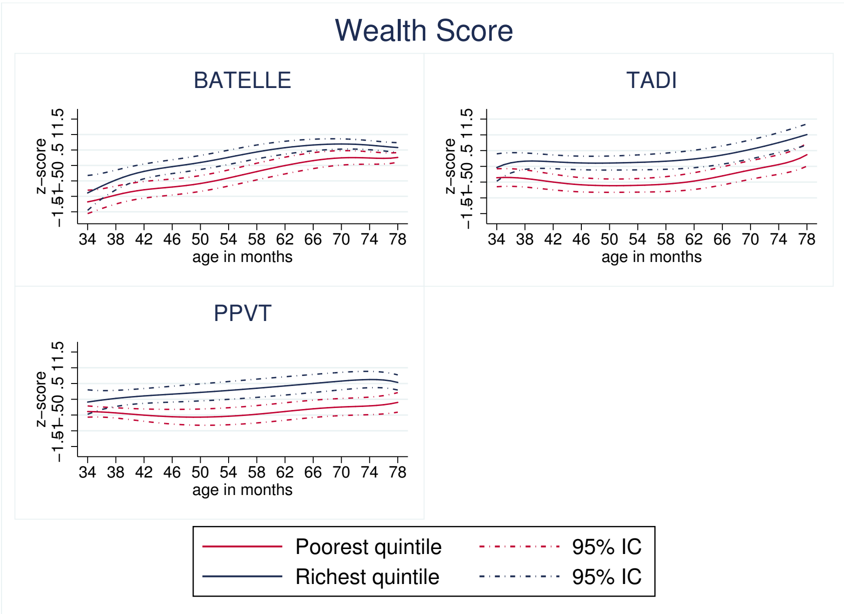


Figure 2.6: Child's Cognitive Measures and Wealth score, ELPI 2012

2.4.2.2 Socioeconomic Status (SES) score

As there is no consensus in which multivariate analysis is the optimal to construct an SES score, I use the PCA method[9] because it does not require multivariate normal distribution assumption. McLoyd (1998) proposes to construct an SES score using home information of education, occupation and income. The education variable is defined as the parental education measure in years of schooling starting in 0 (no education), while the occupational variable is defined as the parental occupational position starting from unpaid workers to employers. Finally, the income variable is defined as the per capita income, in other words, the total familiar income divided by the number of people in the home. For SES, I also construct a score for each sample, where the first principal component explains on average 44% of the variation in these variables (using multiple imputation techniques before generating the factor when there are missing values). The SES factor is placed into quintiles where the first quintile is the poorest. Nevertheless, once the score is constructed (not only by PCA but also by FA) the Kaiser-Meyer-Olkin measure is between 0.50 to 0.59 leading a miserable sampling adequacy. This does not result in some components to be used by the scree plot criterion, which leads to choosing the use of only the wealth score as the measure of socioeconomic status in the empirical analysis. Appendix 2 shows the relevant statistics and figures for screen plot criterion that supports the decision.

2.4.2.3 Home Assessment scores

A crucial determinant of early childhood development is the information that is possible to collect from the child's environment. As shown in Section 2.3, one of the key systems for childhood development is the *Family System* in which parents and other regular caregivers are important assets of the influence of the environment during early childhood as well as household equipment. The most worldwide known instruments are the HOME inventory as well as the FCI. Table A3.1 and Table A3.2 in Appendix 3 present the statistics of the items from the original instruments that are included in the survey and used

[9] Independently of the factor strategy all methods create similar scores. The pairwise correlation between PCA and the two FA is about 0.92-0.99 depending on the sample. Appendix 2 shows these results using the first wave of the survey.

in this analysis for both waves. Regarding the HOME, they are grouped in 2 subscales for the first wave: emotional and verbal responsivity and paternal involvement with the child and in 2 subscales for the second wave: emotional and verbal responsivity, and learning materials. Regarding the FCI, two other measures related to parent-child activities and learning materials in the home were constructed with information included in the first wave and a parent-child activities factor using the second wave[10]. Figure A.0.1 in Appendix 3 show the proportion answering yes for each item.

For creating the home assessment scores, in particular for the HOME inventory, I use a two-stage procedure. Firstly, I use a Factor Analysis initially as exploratory regarding the number of factors that can be extracted from only the HOME inventory as it contains multiples subscales[11]. Two scores are extracted according to the screen plot criterion. Each score is related with each subscale. Table A3.3 in Appendix 3 presents the factor analysis for both waves. For the second step, I use a standard two-parameter IRT model to analyse the extent to which the set of different items for both the HOME and FCI can be used to estimate one or more underlying latent traits representing the quality of the home environment (HE). Note that each of the items j of these inventories are a discrete binary variable. Thus, let define the score $HE_{i,j}^{k,*}$ in the following way:

$$HE_{i,j}^{k,*} = a_j^{HE,k}(\theta_i - b_j^{HE,k}) + u_{i,j}^{HE,k}, \tag{2.2}$$

where $u_{i,j}^{HE,k}$ is the measurement error that has logistic distribution and let θ_i be normally distributed with mean zero and variance σ^2. The parameters $a_j^{HE,k}$ and $b_j^{HE,k}$ capture the difficulty and the informativeness of a given item in each scale k. Let $HE_{i,j}^{k} \in \{0,1\}$ denote the observed score for person i in item j from the HE scale k. It follows that:

$$HE_{i,j}^{k} = \begin{cases} 0, & \text{if } HE_{i,j}^{k,*} \leq 0, \\ 1, & \text{if } HE_{i,j}^{k,*} > 0. \end{cases}$$

[10]Table A3.1 and Table A3.2 show the items of each measure.

[11]Independently of the factor strategy all methods create similar indexes. The pairwise correlation between PCA and FA is about 0.80-0.83 depending on the sample. All variables are standardised. Responses are incorporated to the maximum likelihood extraction method with 50 iterations and with varimax rotation instead of promax due to the correlation.

The estimation is by maximum likelihood and uses GaussHermite quadrature to approximate the log likelihood using 7 points. Letting for each k: $p_{i,j} = Pr(Y_{i,j} = 1 \mid b_j, a_j, \theta_i)$ and $q_{i,j} = 1 - p_{i,j}$, the conditional density for person i is:

$$f(y_i \mid B, \theta_i) = \prod_{j=1}^{J} p_{i,j}^{Y_{i,j}} q_{i,j}^{1-Y_{i,j}}, \tag{2.3}$$

where $y_i = (Y_{1,i}, ..., Y_{J,i})$, $B = (b_1, ..., b_J, a_1, ..., a_J)$.

The log likelihood is the sum of the I individual log likelihoods:

$$\mathcal{L}_i(B) = \int_{-\infty}^{\infty} f(y_i \mid B, \theta_i)\phi(\theta_i)d\theta_i, \tag{2.4}$$

where $\phi(\cdot)$ is the standard normal density function.

The IRT estimation of each scale k are present in Appendix 3 in Tables A.2 to A.7. The scores generated are for *Learning materials*, *Parent-Child Interaction* and *Emotional and Verbal Responsivity* for both waves, and *Paternal Involvement* only for the first wave.

Table 2.1 shows the results for *Parent-Child Interaction* in the second wave. Using the IRT parameterisation described above, the parameter $\alpha_j = a_j$ and $b_j = -\beta_j/\alpha_j$, where the parameter b_j is the difficulty. The item is more difficult the higher it's level of b_j and therefore the lower the level of β_j.

Table 2.1: IRT analysis Act_2012

	(1) TADI	(2) BDI	(3) TVIP	(4) CBCL1
Act_LookBooks				
αi	1.519***	1.522***	1.521***	1.509***
	(0.052)	(0.052)	(0.052)	(0.056)
βi	2.052***	2.049***	2.059***	2.027***
	(0.045)	(0.045)	(0.046)	(0.049)
Act_Stories				
αi	1.678***	1.690***	1.688***	1.678***
	(0.057)	(0.057)	(0.058)	(0.063)
βi	2.313***	2.322***	2.336***	2.356***
	(0.052)	(0.052)	(0.053)	(0.058)
Act_Sing				
αi	1.729***	1.731***	1.727***	1.671***
	(0.066)	(0.066)	(0.066)	(0.072)
βi	3.165***	3.155***	3.174***	3.252***
	(0.070)	(0.070)	(0.071)	(0.078)
Act_GoOut1				
αi	1.013***	1.005***	1.008***	1.030***
	(0.038)	(0.037)	(0.038)	(0.041)
βi	1.121***	1.112***	1.134***	1.145***
	(0.029)	(0.028)	(0.029)	(0.031)
Act_GoOut2				
αi	1.282***	1.282***	1.277***	1.252***
	(0.047)	(0.047)	(0.047)	(0.051)
βi	-0.781***	-0.787***	-0.771***	-0.778***
	(0.029)	(0.029)	(0.029)	(0.031)
Act_TalkingDrawing				
αi	1.812***	1.818***	1.839***	1.783***
	(0.094)	(0.094)	(0.096)	(0.101)
βi	4.657***	4.670***	4.699***	4.643***
	(0.125)	(0.126)	(0.128)	(0.135)
Act_AnimalSounds				
αi	3.994***	3.968***	4.010***	3.805***
	(0.180)	(0.178)	(0.183)	(0.185)
βi	5.525***	5.508***	5.528***	5.508***
	(0.212)	(0.211)	(0.215)	(0.223)
Act_Colours				
αi	5.014***	5.039***	5.112***	4.851***
	(0.261)	(0.263)	(0.271)	(0.274)
βi	6.869***	6.903***	6.974***	6.928***
	(0.320)	(0.323)	(0.333)	(0.345)
Act_Numbers				
αi	8.181***	8.438***	8.382***	7.542***
	(0.865)	(0.944)	(0.931)	(0.769)
βi	12.06***	12.41***	12.36***	11.38***
	(1.189)	(1.299)	(1.282)	(1.074)
Act_Letters				
αi	5.130***	5.175***	5.280***	4.712***
	(0.324)	(0.332)	(0.348)	(0.295)
βi	7.369***	7.429***	7.582***	6.870***
	(0.409)	(0.420)	(0.442)	(0.373)
var(Theta)				
βi	1	1	1	1
	(.)	(.)	(.)	(.)
N	10826	10869	10672	9262

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

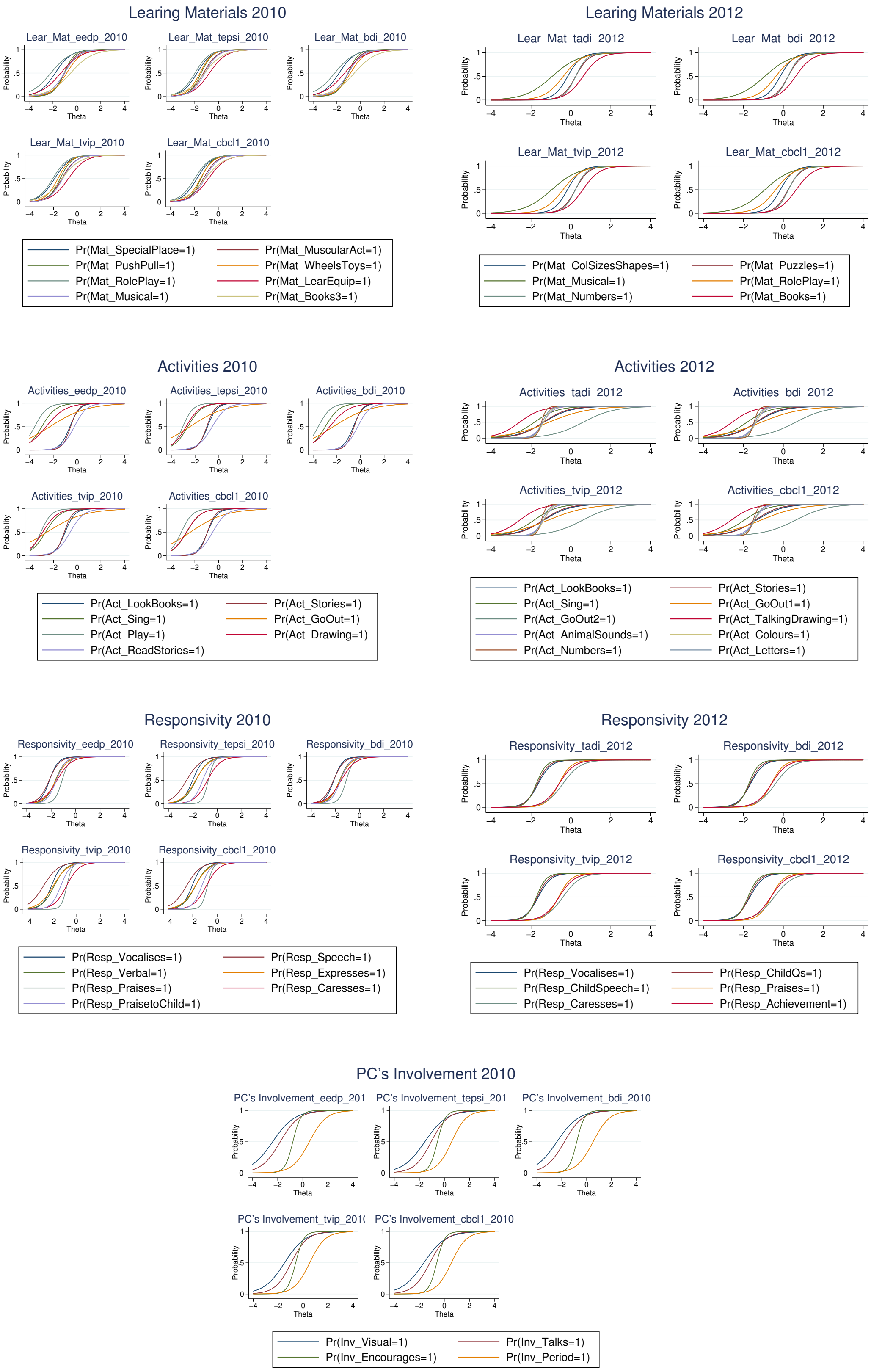


Figure 2.7: Home Assessments: Item Characteristic Curves (ICCs)

In this case, all the items are informative as they have significant and a high $\alpha's$, which corresponds to a steeper slope and should result in a better item fit, and some are more difficult than others but all significant. This is not the same as all the scores where most of them have only loose items but all significant. An informative score would have items that have discriminatory power as well as variability in difficulty, under this type of score one can identify different types of quality environment. Figure 2.7 shows the Item Characteristic Curves (ICCs) for each scale k, most of them show the variation of the ICCs in the negative side of the graph which means that even though they are informative, they are only informative to explain low levels of the quality environment. In this sense, all the results presented in Section 2.5 regarding home assessments are the lower bound.

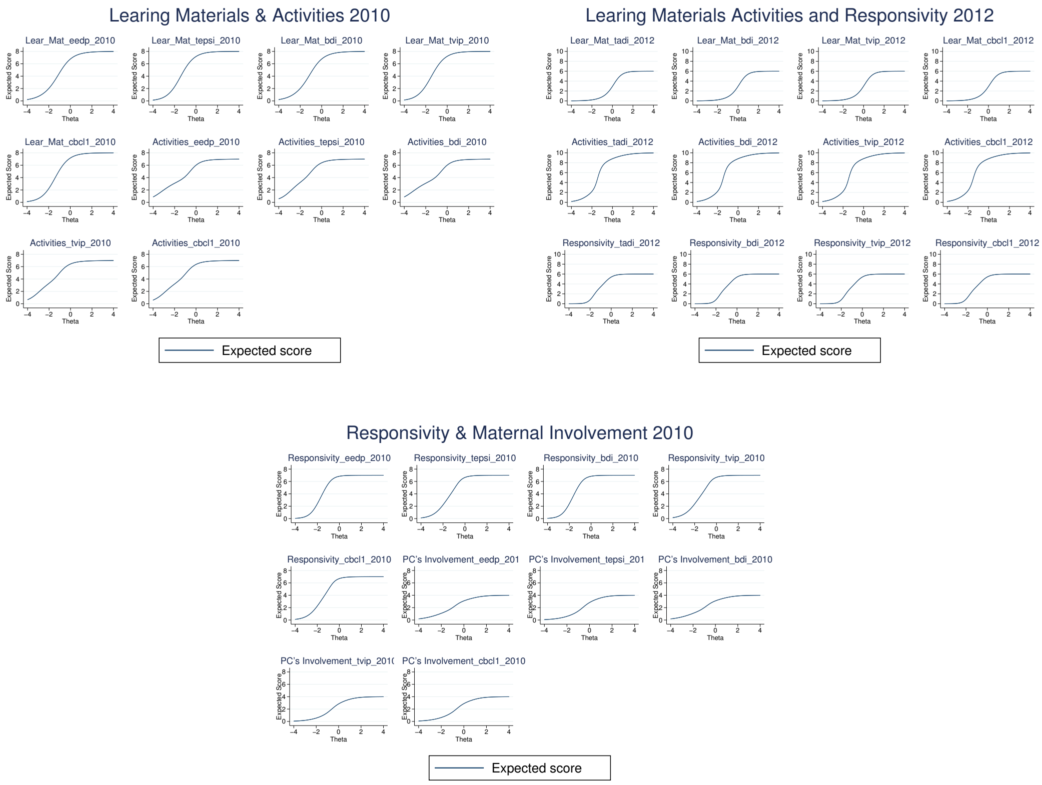


Figure 2.8: Home Assessments: Test Characteristic Curves (TCCs)

Continuing with the IRT models, Figure 2.8 plot the test characteristic curves (TCCs) which are the expected scores for the latent traits for each scale k and Figure 2.9 show the test information functions (TIFs) which are the sum up of all the item information functions that tells us how well the estimation of the latent trait can estimate household locations. The scores provide maximum information for households approximately located at θ =-0.5. As we move away from that point in either direction, the standard error of the TIF increases and the instrument provides less precise information about θ.

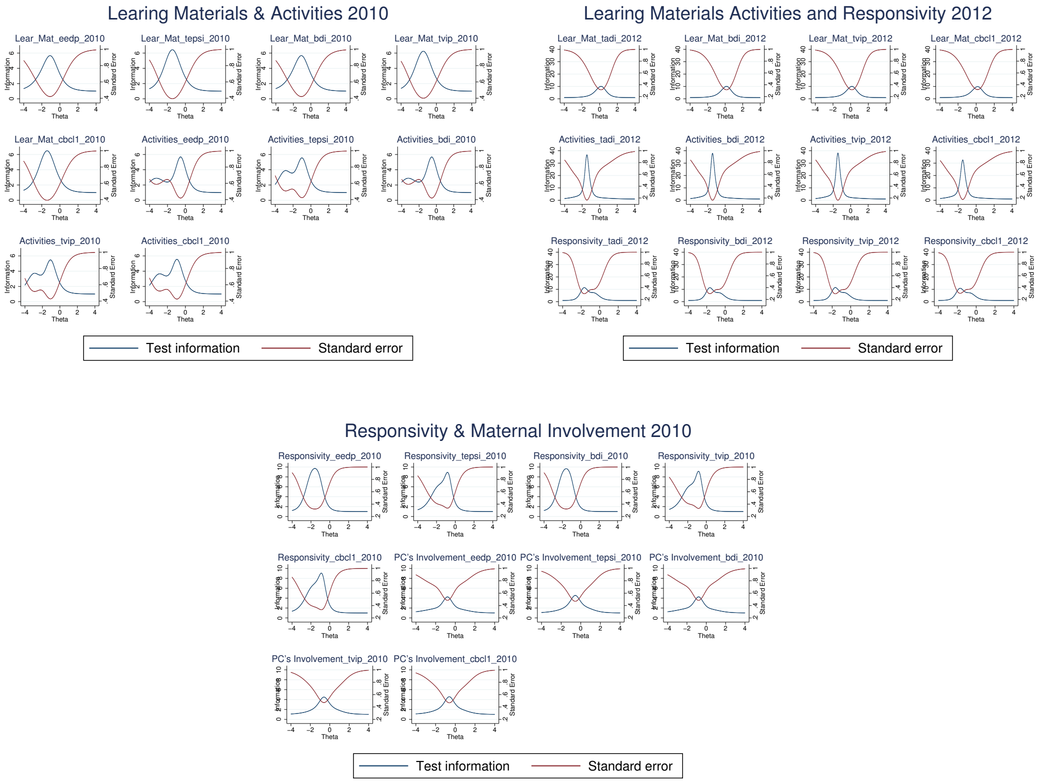


Figure 2.9: Home Assessments: Test Information Functions (TIFs)

2.5 Methodology

Throughout this section, I work out a methodology to disentangle the determinants of cognitive and socioemotional childhood development and hence to assess the impact of public policies that intervene optimally and successfully early childhood development. To achieve this purpose, I first analyse children's cognitive and socioemotional measures for each sample defined by age groups of children. Then, I study the socioeconomic gradients that may exist. Finally, I perform regression analysis for each child's cognitive test.

2.5.1 Socioeconomic Gradients and Childhood Development

The main goal of the first wave of the survey is to assess the cognitive and socioemotional abilities as the physical status of children below 5 years old and their mothers or primary caregivers. The first wave of the ELPI has information about 15,175 households, who have on average 1.1 children under 5 years old, 4.7 people living in the household, an average monthly income per capita equal to 149.92 pounds, 46.6% are male, 40.2% live in the Metropolitan Region (which contains Chile's capital), 35.5% are employed and only 4.2% is composed of a household head aged 45 years old or older. This is because the sampling households for the first wave are those who had a child born on January 1, 2006, and August 31, 2009, so they are mostly households with young household heads.[12]

The first wave of the ELPI has an achievement of 91.6% of measurement assessment, which means, that a total of 13,895 households agreed to be assessed, and as explained in Section 2.4 during the first wave four cognitive and one socioemotional test was applied resulting in five samples that can be used for the empirical analysis (see Table A4 2 in Appendix 4). For the second wave, I use only the panel data and no the information collected in the refreshment sample for children below 3 years old. The number of households from the panel data that were also interviewed in the second wave was 12,898, so the attrition level is approximately 15.0%. During the second wave, three cognitive and one socioemotional test was applied resulting in four samples that can be used for the empirical analysis (see Table A4 3 in Appendix 4). There are one cognitive and one socioemotional test that was applied in both waves, the TVIP and the CBCL1 respectively. Most of the primary caregivers are the mother of the child, in fact, a 98.0% so from now on the mother or primary caregiver will be treated indistinctly as mothers.

As we have several tests for different ages, I divide the sample into two

[12]For details see Table A4.1 and Figure A4.1 in Appendix 4.

early childhood stages: an early stage from 0 to 24 months old, and a later stage from 24/30 to 60 months old in the first wave and until 80 months in the second wave. Figure A4.2 shows the child's age distribution during the first wave. A 37.7% of children are in the early stage.

The evidence from the literature review points out that the critical factor involving children's cognitive and socioemotional development is the socioeconomic level that is constructed with a set of household information. In fact, recent studies in economics, neuroscience and psychology have found that Wealth or Socioeconomic Status (SES) are associated with cognitive and non-cognitive childhood development (Naudeau, 2010; Paxson and Schady, 2005; Heckman and Masterov, 2007; Gagné 2003; Deaton 2003; Schellenberg et al., 2003; Crawford, et al., 2010; Noble et al., 2007 and OECD, 2009 among others).

Section 2.4.2 concludes that the Wealth score is the best measure of the socioeconomic household level after dealing with multivariate analysis. Figure 2.10 shows the socioeconomic gradients for childhood development involving several dimensions: cognition, socioemotional and physical. The socioeconomic gradients statistically account for a portion of the variance in each child's test (calculated as the z-score) with a more significant variances in the later stage. Specifically, the socioeconomic gradient account around 0.4 standard deviation during the first stage of development measured by the EEDP and BDI during the first wave, instead, the socioeconomic gradient account around 0.8 standard deviation during the second stage of development measured by the TEPSI and TVIP during the first wave and the TADI, BDI and TVIP during the second wave.

A key result implies there are socioeconomic gradients in all cognitive tests between the poorest and wealthiest quintiles generated from the socioeconomic gradients. Regarding socioemotional tests, the socioeconomic gradients statistically account for around 0.5 standard deviations measured by the CBCL1 during both waves.

Figure 2.10 also shows the socioeconomic gradients for weight and length at birth which has more lower standard deviation for both, but it is essential to take this into account when analysing the standard deviations for older age-stages after birth.

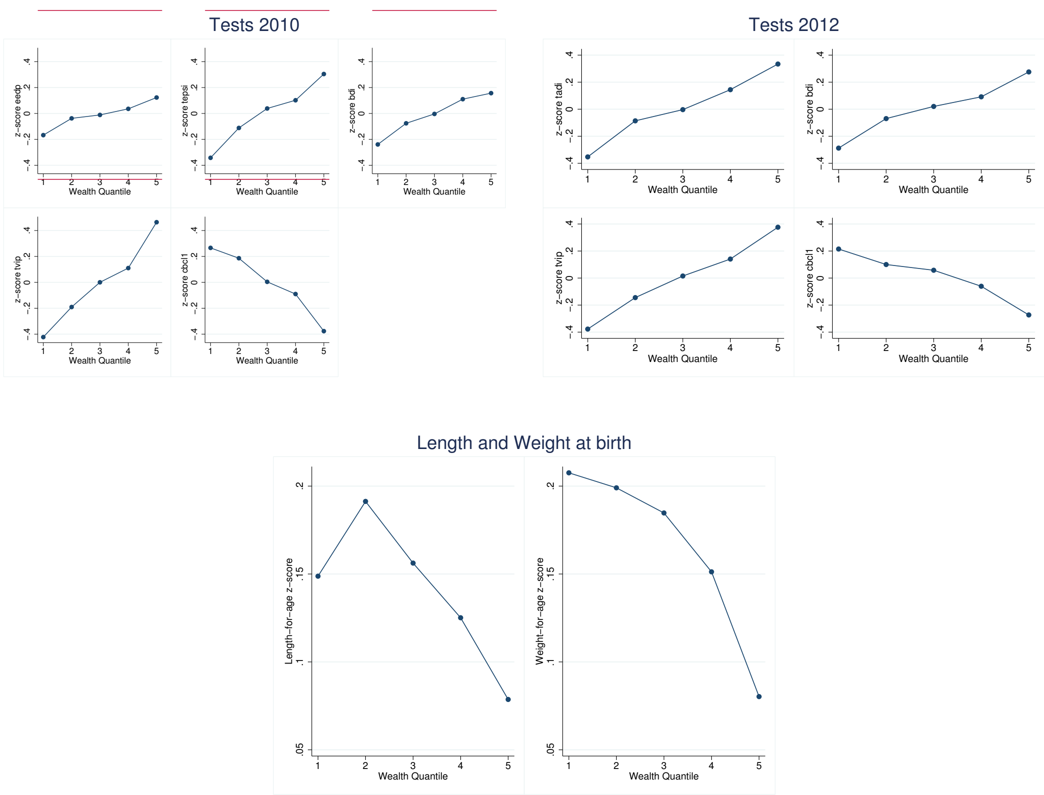


Figure 2.10: Socioeconomic Gradients for Childhood Development

2.5.2 Estimation strategy

Given the presence of socioeconomic gradients for several dimensions of childhood development in both stages, the next step is to spotlight the primary sources that affect children's development. To do so, I use contemporaneous, and value-added production function specifications. The contemporaneous production function for human capital relates cognitive or socioemotional measures to the systems proposed by the Ecological Environment Model for childhood development presented in Section 2.3 which establishes systems that are the basis for flourishing children's abilities. Let's assume that the production function is linear:

$$\begin{aligned} h_{k,t,i} &= \gamma_0 + \gamma_{h0} h_{k,t=0,i} + \gamma_{pc} P^c_{k,t,i} + \gamma_{ps} P^s_{k,t,i} + \gamma_{he} HE_{k,t,i} \\ &+ \gamma_{cc} CC_{k,t,i} + \gamma_x X_{k,t,i} + \mu_{k,t,i}, \end{aligned} \tag{2.5}$$

where $h_{k,t,i}$ denote the cognitive or socioemotional measure $k = c, s$ for child i at age t, the $h_{k,t=0,i}$ vector represent the Child System which is defined by the initial conditions given by genetic, in this case, I use the child's physical information at birth (the z-score for weight and length). For the Family System, we use three vector, the $P^c_{k,t,i}$ which includes maternal education as well as the mother's cognitive test (WAIS), vector $P^s_{k,t,i}$ which includes parental mental health through mother's socioemotional test (BFI) and the vector $HE_{k,t,i}$ which includes all the variables that fall into maternal characteristics and family background categories, in particular, mother's age and employment status as well as indicators for younger or older children in the household and if the family is a two-parent family. Other variables included are the home assessment using the scores described in Section 2.4.2.3 which includes parental style and a measure of household equipment: emotional and verbal responsivity, paternal involvement, learning materials and parent-child activities. The Environment/Neighbourhood is measured using the indicator if the child attends an educational centre that somehow takes into account the child-care or preschool centres access and equipment which is addressed by the vector $CC_{k,t,i}$. Finally, area's type (urban/rural) and the socioeconomic gradient are used as measures of socioeconomics' equity in the Society System through the vector $X_{k,t,i}$ [13].

In the estimation procedure, there are some difficulties that need to be addressed, firstly, we do not observe $h_{k,t}$, P_c, P^c, P^s and HE_t directly. Indeed, most of the measures used to create factors for cognitive, socioemotional and home assessments (investment) inputs are measured with error. One way to address the measurement error is the use of IRT analysis for the estimation of unbiased investments parameters. Another issue is that problems of multicollinearity and endogeneity may arise for this I assume that all the omitted factors are orthogonal to the included input measures. A second way to address these issues is to use a value-added specification (Todd and Wolpin, 2003, 2006) as the following:

[13] Table A4 2 and A4 3 show descriptive analysis of the sample of children depending on the applied test for the first and second wave.

$$\begin{aligned} h_{k,t,i} &= \gamma_0 + \gamma_k h_{k,t-1,i} + \gamma_{h0} h_{k,t=0,i} + \gamma_{pc} P^c_{k,t,i} + \gamma_{ps} P^s_{k,t,i} \\ &+ \gamma_{he} HE_{k,t,i} + \gamma_{he} HE_{k,t-1,i} + \gamma_{cc} CC_{k,t,i} + \gamma_x X_{k,t,i} + \mu_{k,t,i}, \quad (2.6) \end{aligned}$$

Nine models are estimated using specification (2.5) and two using specification (2.6) due to the availability of lagged information for one cognitive and one socioemotional test, the TVIP and CVCL1 respectively as well as lagged information for HE.

2.5.3 Results

Tables A.8 to A.16 in Appendix 4 present the contemporaneous linear regression models using ordinary least squares and where the dependent variable is the children's cognitive or socioemotional test z-score in each wave. Tables A.8 to A.12 in Appendix 4 are for the first wave meanwhile tables A.13 to A.16 are for the second wave. Note that the contemporaneous linear regression models are also for different age stage. The early (later) stage is presented in Tables A.8 (EEDP) and A.9 (BDI 2010) (Tables A.10 (TEPSI) to A.16 (CBCL1 2012)) in Appendix 4 [14]. In each estimation, I obtain the socioeconomic gradient in children's test scores. This allows me to confirm the results found it in the Section 2.5.1 about the existence of socioeconomic gradients (specification 1) and successively add to the model estimation demographics, mother's characteristics and other family background (specification 2), child's physical (specification 3), mother's cognitive and socioemotional tests (specification 4) and child's home environment (specification 5).

In the early stage specifications, Table A.8 show that children in the fifth quintile have higher average EEDP z-scores (difference of 0.276 standard deviations) than children in the bottom quintile statistically at 1%. Hence, a socioeconomic gradient in cognitive, psychomotor development is present. However, this gradient begins to disappear and to lose significance since the specification is controlling by other variables like $h_{k,t=0,i}$, $X_{k,t,i}$, and $CC_{k,t,i}$. In general, the

[14] Regression check and deal with multicollinearity, nonlinearities and heteroskedasticity. Also included dummies variable created for each explanatory variable with missing values (coded as a one if the related explanatory variable is missing and zero otherwise).

children's age (in months) and being male have positive and negative significant effects on EEDP respectively, mother's age and if the child attends to a child-care are significant and negative suggesting that younger children are not translating what they receive in the child-care to improvement in the EEDP. The main result is the drop in the socioeconomic gradient due to the inclusion of child's physical information at birth (weight), mother's socioemotional tests (BFI_Openness), and the parent-child, involvement and responsibility scores which have significantly big and positive effects on EEDP z-scores. In particular, both involvement and responsibility one standard deviation of the scores increase the EEDP z-scores by 0.246 standard deviation. On the other hand, Table A.9 show that the BDI has a z-score that is 0.39 standard deviation higher for children in the bottom quintile, which finished with a z score equal to 0.057 and losing significance when controlling for all variables (specification 5). The latter implies that a bigger socioeconomic gradient is present respect to EEDP due to the results' stability in this test. Child's age and being male have positive and negative significant effects on EEDP respectively. The fall in socioeconomic gradient is caused by the inclusion of physical information at birth (weight), urban household, mother's age and cognition (only verbal WAIS), socioemotional tests (BFI_Openness), and by the learning materials, parent-child, involvement and responsibility scores, which have positive effects on child's general development measured by the BDI. For both, the EEDP and BDI, maternal employment status does not affect.

In the later stage specification, Table A.10 show the TEPSI has a z score that is 0.649 standard deviations higher when compared to children in the bottom quintile, who have a 0.213 z-score without losing any significance after controlling for all variables by decreasing their effectiveness. In general, the children's age (in months) is not significant for most of the specifications and being male have adverse and significant effects on TEPSI, mother's age and if the child attends to preschool are significant and positive suggesting a complementary effect between preschools and household environment, as the learning materials, parent-child, involvement and responsibility scores, have positive effects on child's general development measured by the TEPSI. TableA.11 show that children in the fifth quintile have higher average TVIP z-scores (difference of 0.912 standard deviations) than children in the bottom quintile. This stage has a bigger impact than the early stage in cognition. An important feature is that the effect does not disappear and it is significant, while the specification controls by other variables, resulting in a difference of 0.395. In all specifications, the child's age and if the household is urban have positive and significant effects. Household environment, as the learning materials, parent-child, involvement and responsibility scores, have positive results on child's language measured by the TVIP as well as the mother's cognitive level in both digit span but also language and socioemotional tests (BFI_Extroversion). For both the TEPSI and TVIP, now during the later stage mother's education, age and employment status are significant and positive meanwhile more children in the household are significant but negative effect on TEPSI and TVIP nonetheless having a two-parent family has a positive effect.

For the second wave, Table A.13, Table A.14 and Table A.15 show similar results and reinforcing the results found for the first wave in the later stage specification. Children in the fifth quintile have higher average z-scores than children in the bottom quintile, and this effect does not disappear, and it is significant, while the specifications control for other variables. If the child attends to a preschool this has positive effect in all the measures, mother's characteristics are more important than in early stage in terms of education, cognition and some subscales of socioemotional test are still important, and the home environment is still significant and positive but now the estimation for learning materials is bigger than for early stages.

Regarding the socioemotional scale[15], for both waves, the results are similar, there are socioeconomic gradients that are still significant after controlling for all the variables. If the child is male, have a negative and significant effect as well if they attend to preschool. Mother's education and age have positive and significant effect meanwhile having younger children in the household have a negative and significant effect. Having both parents have a positive effect as well as child's weight at birth and the mother's cognition level. For the first time, all the sub scales of the mother's socioemotional test are (positively) correlated with the child's socioemotional skills (before, only some of them were correlated with chid's cognition skills) for the first wave and only three (BFI_Agreeableness, BFI_Conscientiousness and BFI_Neuroticism) for the sec-

[15]note that positive z-score of the CBCL1 means poor socioemotional skills

ond wave. The home environment has also positive and significant effect.

Tables A.17 and A.18 present the added value specifications using ordinary least squares and where the dependent variable is the children's cognitive or socioemotional test z-score present in both waves. The lagged measure is the one that presents the bigger effect on the TVIP or CBCL1 respectively. For the TVIP, there is socioeconomic gradient instead for the CBCL1 this one disappears completely. For both measures, the mother's cognition (measured by the WAIS vocabulary) is significant and positive but very small, and the mother's age and education are significant and positive. For the TVIP, learning materials in t and parent-child interaction in $t-1$ have a positive and significant but small effect on the language child's skill. For the CBCL1, learning materials in t and emotional and verbal responsivity in t and $t-1$ have a positive and significant effect on the socioemotional child's skill as well as some subscales for mother's socioemotional development (BFI_Agreeableness and BFI_Neuroticism).

Note that all the results regarding the HE variables are the lower bound effect due to the fact that most of them show variation in the negative side of the scores which means that even though they are informative, they are only informative to explain low levels of the quality environment.

The analysis of the empirical evidence, using a contemporaneous production function of human capital, suggests that all tests present socioeconomic gradients in child development when no other controls are considered. However, once controlling, the gradient starts to decrease and in the worst case to lose significance for the early stage. In general, the fall in the socioeconomic gradient is explained by the inclusion of physical information at birth, mother's cognitive and socioemotional tests and the home environment scores, all of them having positive effects on the child's cognitive and socioemotional development. Once that the value-added estimation is used, the lagged of the cognitive or socioemotional measure, TVIP or CBCL1 respectively presents the more significant effect on the outcome variable.

2.6 Concluding remarks

In this chapter, I contributed to the discussion about the determinants of early childhood cognitive development through theoretical and empirical evidence. This chapter is one of the few studies that deal with this challenging concern in Chile and developing countries, mainly due to the inadequate information about the topic. I take advantage of a longitudinal survey designed to characterise children development, using its first two waves and that contains a set of cognitive, socioemotional and anthropometrics measures for both the children and their mothers that help to fill the emptiness of such evidence due to the typical unobserved parental abilities issue, measurement error and endogeneity, which allows a proper monitoring of early childhood.

The determinants of early childhood development, particularly, cognitive and non-cognitive skills, are studied through the estimation of contemporaneous and value-added cognitive and non-cognitive production functions, as well as the use of factor analysis such as item response theory for reducing the number of inputs using a theoretical framework.

I show that when the socioeconomic wealth score is included as a whole, the overall association with cognitive and socioemotional child's development can be obtained for early and later stages. This is also supported by empirical analysis indicating significant socioeconomic gradients in all cognitive and socioemotional tests between the poorest and richest quintiles. The latter implies that disadvantaged children are liable. Once controlling by other observables, the gradient starts to decrease and in some cases to lose significance; there is a significant effect of mother's characteristics and family background at later stage development (above 24/30 months old) measured principally by mother's education, age and cognitive skills, if the family is a two-parent family, the presence of younger/older children as well as home environment measures by parent-child activities, learning materials, parental involvement and verbal and emotional responsibility scores. The later stage development also adds a significant effect on attending a preschool. The previous determinants drive the fall in the socioeconomic gradient in both stages.

Regarding the non-cognitive skills, for both waves, the results are similar, there are socioeconomic gradients that are still significant after controlling for all the variables. If the child is male, have an adverse and significant effect as well if they attend to preschool. Mother's education and age have positive and significant effect meanwhile having younger children in the household have a negative and significant effect. Having both parents have a positive effect as well as child's weight at birth and the mother's cognition level. For the first time, all the subscales of the mother's socioemotional test are (positively) correlated with the child's socioemotional skills. The home environment continues presenting positive and significant effect on child's development.

A common limitation of previous studies and also this one is that they have failed to control for potential endogeneity problem fully: parents that work may be systematically different from parents who do not work, and the child's skills itself may influence parental decisions of whether to work or not, moreover, parents are heterogeneous in their skill endowments, the constraints they face, and their tastes, therefore is crucial to allow parental decisions to depend on these unobserved heterogeneous characteristics of both parents, as well as, measurement error problem. For this reason, in the next chapters, I try to deal with these issues including a measurement error system in the estimation of the production function for human capital and estimating a model of parental investment choices jointly with a child production function for cognitive and socioemotional abilities.

These results raise important questions about the relevance about quantity v/s quality evaluation, about the ideal form of early interventionism, about the implications on the optimal design, and about the implementation of an optimal and successful public policy designed to improve child development, for example, between the optimal time for mother to enter or to go back to the labour market as well as when to send the children to child-care centres or preschool institutions.

Chapter 3

The myths behind the Technology of Human Capital Formation in Childhood

3.1 Introduction

Recent research spotlights the effect of early influences and investments by parents during early childhood on brain blooming, learning skills and non-cognitive abilities suggesting that the first years of life are a crucial period for children's development. Non-cognitive abilities are associated with patience, perseverance, temperament, motivation, self-control and self-esteem among others. Heckman et al. (2006) find that cognitive and non-cognitive abilities are critical to the future performance on a series of social and labour market outcomes: enrolment rates, wages, work experience, crime rates, early pregnancies, drug use in the labour market, among others providing evidence of the effect of cognitive and non-cognitive abilities in socioeconomic success. Nonetheless, gaps in both cognitive and non-cognitive skills emerge at early ages before preschool and remain constant in the life-cycle; consequently, a clear comprehension of the different channels for the human capital formation process in childhood is essential for improving childhood development. Only when these channels, mostly related to genetics, investments and environment, are adequately incorporated in the study of the human capital formation will be possible to tackle early gaps in childhood and formulating efficient public policies.

Children growing in poverty, either in developed or developing countries, accumulate substantial lags regarding their development during these first years. Investments and environment seem to be crucial determinant factors, in fact, Cunha et al. (2010) find that early deficits are difficult to reverse and early investments can yield high returns hence reducing gaps in multiple dimensions at early stages contributes to shrink social inequality, poverty and exclusion, and to foster economic development. During the last years, research has been done in terms of implementing early childhood programs for improving their abilities based on the results of three crucial randomised controlled trials: Jamaican home-visiting program (Grantham-McGregor et al. 1991), Perry Preschool Program (Schweinhart et al. 2005) and Abecedarian (Campbell et al. 2002). The fundamental policy questions are what is the causal effect of an early childhood program over time relative to a particular alternative? Whether or not the program should be subsidised by governments and scale them?

Despite these recent advances, there is still very little known about the return to cognitive and non-cognitive skills in developing countries. Most of these countries do not have the resources for running well implemented Randomized Controlled Trials to study the effects of an early childhood program. In order to shed light on these questions, this chapter use a dynamic structural approach to understand the mechanisms behind observed decisions in the human capital formation in childhood in a way that is consistent with accepted theories of economic behaviour and hence to identify the factors behind the early gaps and provide feedback to policymakers so better policies are implemented to narrow these gaps using longitudinal data for Chile.

Initially the studies for human capital formation were based on one single period and inputs in the production function assumed to be perfect substitutes (Becker (1964), Mincer (1974), Aiyagari et al. (2002), Loiter (1992, 1997)), as a result of the importance of disentangling the process behind human capital formation and richer longitudinal data available in early childhood more complex model were developed. An improve methodology in terms of the conventional approach to estimating cognitive production functions was proposed by Todd and Wolpin, 2003, 2006 offering a value-added "plus" specification with lagged input variables which is consistent with theoretical notions that human capital formation is a cumulative process, but this approach has not addressed the multiplicity of parental inputs, measurement error in the inputs as well as the fact that inputs may be chosen endogenously with respect to child's endowments and hence the mechanisms through which parental behaviour could affect their children outcome. Cunha et al. (2010) develop a dynamic factor model that exploits covariance equations restrictions to secure identification of a multistage technology for child investment, in particular, include multiple-stages to match patterns of multiple returns, inputs can be substitutes or complements, different age-stages and parental investments can be self-productive. The advantage of this methodology is that it uses dynamic factor models to solve the problem of endogeneity of inputs and the multiplicity of inputs relative to measures. Attanasio et al. (2015b) follows a similar approach as Cunha et al. (2010) to model the accumulation of future skills as a process that is determined by the child's current stock of skills, parents' investments and parental human capital for this they develop an estimation approach that combines maximum likelihood estimation using the EM algorithm with simulation methods and the control function approach to account for the endogeneity of investments. The advantage of this approach is that it is both flexible and a tractable estimation technique to use but at the same time does not allow for a joint estimation of the measurement system and the production functions.

The model I estimate in this chapter uses the framework developed by Cunha et al. (2010) for the estimation of dynamic and multidimensional latent factor model of skills accounting for the endogeneity of investment and dealing with measurement error. One contribution to the literature that I add to Cunha et al. (2010) by including a multi-dimensional parental investment vector not only regarding material resources (monetary investment) and quality time investment but also regarding cognitive stimulation and emotional support. This model provides two key parameters, the self-productivity of skills (if the child learn how to count, then he can use it to learn other skills which means that skills are self-reinforcing and persist into future periods) and dynamic complementarity (synergy of investments at different t), hence, a second contribution is to analyse if complementarities change with age in a context of a developing country. A third contribution is concerning the optimal number of dimensions, are two dimensions enough or should we expand the system to three or four (cognition, socio-emotional, language, health, etc.)?

The estimation of these technologies of human capital formation in childhood allows answering appropriately what investment practices or strategies are more efficient in the creation of cognitive and non-cognitive skills (and other dimensions).

The previous chapter uses the methodology proposed by Todd and Wolpin, 2003, 2006 and finds that mother's skills, the home environment as well as attending to preschool are the most important determinants of skill formation in childhood. This chapter attempts to contribute to the growing early childhood literature through addressing some of the myths behind the estimation of the technology of human capital and the real returns on parental investment for different stages during childhood.

This chapter also contributes from previous research as include a rich Chilean data to apply the state-of-the-art methodology in the estimation of the production function. I use data from a recently longitudinal survey designed in Chile to characterise children development, using its first two waves, which were collected in 2010 and 2012, and that contains a set of cognitive, socioemotional and anthropometrics measures for both the children and the mothers or primary caregivers. There is also information about home assessment using the Home Observation for Measurement of the Environment (HOME) inventory and the Family Care Indicators (FCI) that help to measure intermediate outcomes and mediators for early childhood development. This information deals with the typical unobserved parental abilities issue, which allows a proper monitoring of early childhood. Chile is one of the major industrialised countries in Latin America, but at the same time, is the country with higher inequalities among its population regarding income relative to other OECD countries (OECD, 2011). This fact leads perhaps too strong socioeconomic gradients by cognitive and non-cognitive development.

This chapter proceeds as follows: Section 2 presents the framework of parental investments and the technology of human capital in childhood and how is possible to deal with measurement error and endogeneity of inputs.

Section 3 presents the survey data and the multiple empirical specifications for the estimation of the technology of human capital in childhood. Section 4 presents the main results. Section 5 concludes.

3.2 Estimation of the Technology of Human Capital Formation

There is growing recognition that multiple skills are essential predictors and likely determinants of success in many aspects of life. Although a variety of methods are used to measure these skills, there is no agreement on the best ways to do so. Parental environments and investments at different ages and stages of childhood development determine skills. Recent studies have demonstrated how multiple factors relate in a complicated way (Cuhna et al. (2007, 2010)). Understanding the factors affecting the evolution of multiple skills is crucial for the design of effective public policies in developing countries. As pointed out by Cuhna et al. (2010), it is necessary to estimate a multistage technology to capture different development phases in the life cycle of a child. For this, they identified a more general nonlinear technology by extending linear state space and factor analysis to a nonlinear setting. This allows eliminating the assumption that early and late investments are perfect substitutes over the feasible set of inputs.

To estimate the technology of human capital, I use the two-step procedure proposed by Cuhna et al. (2010). In this technology, inputs at each t produce outputs at $t+1$ and children's development is driven by a certain number of latent factors which are reflected in measurements. Similarly, investment is also driven home environmental measurements. In the first step, as some issues need to be addressed, such as the presence of endogeneity (correlation with the unobserved shock) and measurement error in the data for the inputs in the technology, I estimate the measurement model of the latent factors based on a nonlinear dynamic factor model and exploit cross-equation restrictions (covariance restrictions) proving that I can identify all of them. In the second step, I estimate together with the dynamic factor model and the non-linear technologies. I consider initially two technologies, one for the production of cognitive skills and one for the production of non-cognitive skills. One innovation in the estimation is to add a new multi-dimensional parental investment, specifically, measured as material resources (monetary investment) and quality time investment as well as cognitive stimulation and socioemotional support. A second possible innovation in the estimation is to increase the set of technologies including also a third and even fourth skill, for example, health or language[1].

I estimate different specifications of the technology of skill formation for cognitive and non-cognitive skills in childhood, for all of them the production function is assumed to be the Constant Elasticity of Substitution (CES) to allow for complementarities of inputs. *Specification 1* assumes one input for measuring investment for only one stage of child's development, *Specification 2* estimates previous specification but for multiple stages of child's development, there are 3 child age stages at $t = 0$ and hence 3 more at $t = 1$, each of them accounting for a different childhood stage, the gap between $t = 0$ and $t = 1$ is approximately 20 months. The first age stage is for children aged 7-23 months, the second one for children aged 24-47 months and the last one for children aged 48-58 months. The decision behind the stages is mostly due to the different skills that are measured at each stage. *Specification 3* assumes multi-dimensional parental investment, specifically, measured as material resources (monetary investment) and quality time investment and *Specification 4* assumes multi-dimensional parental investment, specifically, measured as cognitive stimulation and socioemotional support.

The explanation about the procedure behind the estimation of the technology of skill formation is based in *Specification 1*, the results hold for the other specifications as well. *Specification 1* incorporates only one input for measuring investment, parental investment i_t in child skills at age t, that accounts for several measures, in particular, parent-child activities, learning materials, emotional and verbal responsivity, paternal involvement, acceptance, child's discipline and activities related with the program *Chile Grows with You*. The program (Vega, 2011) provides free access to childcare centres and parenting advice for children among the poorest 60% and between3-months-old to

[1]I'm currently in the estimation process of a system of three skills which includes cognition, health and non-cognitive skills.

4-years-old. $h_{c,t}$ denotes cognitive skills of the child at age t, $h_{n,t}$ denotes non-cognitive skills of the child at age t. P_c and P_n represent maternal cognitive and non-cognitive skills respectively, A is the total factor productivity and η_t are shocks and/or unmeasurable inputs like the use of childcare of preschool. The technology for skill $k = c, nc$ at period t (where $t = 24 - 47$ months) is:

$$\begin{aligned} h_{k,t+1} &= A_k e^{\eta_t} [\gamma_{k,c} h_{c,t}^{\phi_{c,k}} + \gamma_{k,nc} h_{nc,t}^{\phi_{nc,k}} + \gamma_{k,pc} P_c^{\phi_{pc,k}} \\ &+ \gamma_{k,pnc} P_{nc}^{\phi_{pnc,k}} + \gamma_{k,i} i_t^{\phi_{i,k}}]^{\frac{1}{\phi_k}}, \end{aligned} \tag{3.1}$$

where $\gamma_{k,m} \in [0,1]$, $\sum_m \gamma_{k,m} = 1$ for $m \in [c, nc, pc, pn, i]$. η_t is assumed to be normally distributed and serially independent over all t and includes unobserved shocks or omitted inputs like the choice of childcare or preschools. The elasticity of substitution is $1/(1-\phi_k)$ and measures the level of substitution in the inputs, where $\phi_k \in (-\infty, 1]$, if $\phi_k \to 1$ then the inputs of the production function become perfect substitutes and if $\phi_k \to -\infty$ they become instead perfect complements. The technology has the regular properties: is monotone increasing in its arguments, twice continuously differentiable, and concave in investment input. This formulation assumes that the formation of skills depend on initial conditions, A_k, the stock of skills in period t, parental investment at t, i_t, maternal skills, P_k, and shocks in period t, η_t.

The term A_k which captures the total factor productivity depends on initial conditions that shifts the production function which are potential characteristics that determines the *initial* level of skills. The covariates that enter into the total factor productivity are child's gender, X_g, the birth order, X_b, which is fix and do not to change between t and $t+1$ as in the sample only a 5% gave birth to a new child, the birth height, X_{bh}, and birth weight, X_{bw}. Instead of include these parameters as factors in the production function I assume that these only produce a different starting point (the constant term) in the outcome as they are determining initial levels of skills for these different *types* of children and hence produce direct effects on the outcomes and not through the elasticities.

$$A_k = f(X_g, X_b, X_{bh}, X_{bw}), \tag{3.2}$$

As mentioned before, this model provides two key parameters, the self-productivity of skills and a dynamic complementarity. The self-productivity which includes own and cross effects, is the change of the technology when we have a change in the child's skill:

$$\frac{\partial h_{k,t+1}}{\partial h_{k,t}} > 0$$

The dynamic complementarity measures when stocks of skills acquired by the previous period make investment in current period more productive:

$$\frac{\partial h^2_{k,t+1}}{\partial i_t h_{k,t}} > 0$$

For estimating the parameters of technology (3.16) we need to deal with two problems:

- We do not observe directly skills (children's skills and maternal cognitive P_c and non-cognitive P_{nc} skills) and parental investment i_t;
- Endogeneity of inputs: Parental investment, i_t, is chosen based on the information from η_t which is not observed by the econometrician and which includes unobserved shocks or inputs.

3.2.1 Measurement error

When we do not observe directly children's skills, maternal skills and parental investment, instead, we can use observed proxy. Indeed, most of the measures used to create factors for cognitive, non-cognitive and investments inputs are measured with error. Let's use $h_{k,t}$ to explain the procedure, in this case, I can use measurements $y_{k,t,j}$ where each measurement j (e.g. test scores for $k = c$) are additively separable functions in the (log)[2] of the underlaying trait $h_{k,t}$ and has its own informative factor loading, $\alpha_{k,t,j}$ as following:

[2]the use of natural logarithms keeps the latent factor only take positive values.

$$y_{k,t,j} = \mu_{k,t,j} + \alpha_{k,t,j} ln(h_{k,t}) + \varepsilon_{k,t,j}, \quad (3.3)$$

where $\mu_{k,t,j}$ is the mean and the vector $\varepsilon_{k,t,j}$ captures measurement error. For each measurement system, the goal is to recover the latent factor $h_{k,t}$ which is error-free. Several assumptions are needed for the identification of the measurement system where the covariance restrictions in measurement error are crucial:

1. Define a factor scale $\alpha_{k,t,1} = 1$ in each system, which means we need to standardise the factor loading of one measurement per latent factor equal to 1;
2. Define the factor location by setting $E(ln(h_{k,t})) = 0$;
3. $\varepsilon_{k,t,j}$ are uncorrelated across all measurements;
4. $E(\varepsilon_{k,t,j}) = 0$; and
5. There are at least 3 measures for each factor. The identification draws from the Kotlarsky theorem that states that with two independent measurements per factor, the distribution of the underlaying trait and the measurement error can be identified non-parametrically up to a change of location.

Using these assumptions it is straightforward to proof the identification for the children's skills, maternal skills and parental investment latent factors. Let's use $h_{c,t}$ as an example. I proceed in several steps, first let's consider the following fourth measurements for the cognitive latent factor at t:

$$\begin{aligned}
y_{c,t,1} &= \mu_{c,t,1} + ln(h_{c,t}) + \varepsilon_{c,t,1} \\
y_{c,t,2} &= \mu_{c,t,2} + \alpha_{c,t,2} ln(h_{c,t}) + \varepsilon_{c,t,2} \\
y_{c,t,3} &= \mu_{c,t,3} + \alpha_{c,t,3} ln(h_{c,t}) + \varepsilon_{c,t,3} \\
y_{c,t,4} &= \mu_{c,t,4} + \alpha_{c,t,4} ln(h_{c,t}) + \varepsilon_{c,t,4},
\end{aligned}$$

Initially, we can identify the factor mean for the first measurement: $\mu_{c,t,1} = E[y_{c,t,1}]$ as $E(ln(h_{k,t})) = 0$ and $E(\varepsilon_{k,t,j}) = 0$. The second step is to identify the factor loadings for the measurement system, for this, first I calculate de covariance using assumptions 3 and 4 for the (3.3), (3.4) and (3.6), relationship (3.5) combines results from (3.3) and (3.4) and relationship (3.7) combines results from (3.4) and (3.6):

$$
\begin{aligned}
cov(y_{c,t,1}, y_{c,t,2}) &= \alpha_{c,t,2} var(ln(h_{c,t})) && (3.4)\\
cov(y_{c,t,1}, y_{c,t,3}) &= \alpha_{c,t,3} var(ln(h_{c,t})) && (3.5)\\
\alpha_{c,t,2} &= \alpha_{c,t,3}[cov(y_{c,t,1}, y_{c,t,2})/cov(y_{c,t,1}, y_{c,t,3})] && (3.6)\\
cov(y_{c,t,2}, y_{c,t,3}) &= \alpha_{c,t,2}\alpha_{c,t,3} var(ln(h_{c,t})) && (3.7)\\
&= \alpha_{c,t,3} cov(y_{c,t,1}, y_{c,t,2}), && (3.8)
\end{aligned}
$$

Finally, rearranging I can (3.7) and solving for $\alpha_{c,t,3}$ I can then identify all the factor loadings of the measurement system:

$$
\begin{aligned}
\alpha_{c,t,1} &= 1\\
\alpha_{c,t,3} &= cov(y_{c,t,2}, y_{c,t,3})/cov(y_{c,t,1}, y_{c,t,2})\\
\alpha_{c,t,2} &= cov(y_{c,t,2}, y_{c,t,3})/cov(y_{c,t,1}, y_{c,t,3})\\
\alpha_{c,t,4} &= cov(y_{c,t,2}, y_{c,t,4})/cov(y_{c,t,1}, y_{c,t,2}),
\end{aligned}
$$

Then, the variance of the underlaying trait is derived with the previous results and defined as:

$$
\begin{aligned}
var(h_{c,t}) &= [cov(y_{c,t,1}, y_{c,t,2})/cov(y_{c,t,2}, y_{c,t,3}]cov(y_{c,t,1}, y_{c,t,3})\\
var(\varepsilon_{c,t,1}) &= var(y_{c,t,1}) - var(\widehat{ln(h_{c,t}})) \\
var(\varepsilon_{c,t,2}) &= var(y_{c,t,2}) - \alpha^2_{c,t,2} var(\widehat{ln(h_{c,t}})) \\
var(\varepsilon_{c,t,3}) &= var(y_{c,t,3}) - \alpha^2_{c,t,3} var(\widehat{ln(h_{c,t}})) \\
var(\varepsilon_{c,t,4}) &= var(y_{c,t,4}) - \alpha^2_{c,t,4} var(\widehat{ln(h_{c,t}}))
\end{aligned}
$$

I can follow the same logic in (3.3) and assumptions for identification of the other children's skills, maternal skills and parental investment.

3.2.2 Endogeneity of Inputs

The use of Instruments Variables or within-child/family fixed effect estimators have been implemented for addressing the problem. For the later, it is needed to have multiple observations for the child at different ages or assume that children in the same family have a common heritable component. As explained before, this problem arises because parents adjust their investment decisions to unobserved shocks or inputs that are not observed by the econometrician. Cuhna et al. (2010) propose a solution involving exclusion restrictions based on economic theory in the spirit of control function procedure. Hence, for identification of the model I need some exclusion restrictions, at least as many as endogenous variables are in the model, that determines parental investment decisions but not to be included in the technology of skills.

To solve this I use therefore exclusion restrictions, say z_t, that is contained in the state variables at period t but the reverse is not true, meaning, z_t it is not contained in the children's skill, maternal skills and an invariant-time heterogeneity component, π observed by parents before making investment decisions but they exogenously affect the household budget constraint. z_t contains average comuna-level female (F_W) and male (M_W) wages as they change the household budget constraint at t and hence parental investment decisions, but they are not included in the production function estimation. At least one exclusion restriction is needed to identify the model and thus the parental investment decision but for the case of multi-dimensional investment at least 2 exclusion restrictions will be needed, for this case, I also add to the set the average comuna-level unemployment (U) and the average region-level of investment prices (I_P)[3]. The variable π is assumed to be distributed independently among children and observed by the parent before making investment decisions and hence observed by the econometrician. The log-linear investment policy function, for the case I use two exclusion restrictions, is defined as:

[3]All the variables in z_t are assumed to be measured without error.

$$
\begin{aligned}
ln(i_t) &= \kappa_1 ln(h_{c,t}) + \kappa_2 ln(h_{nc,t}) + \kappa_3 ln(P_c) + \kappa_4 ln(P_{nc}) + \kappa_5 ln(F_{Wt}) \\
&+ \kappa_6 ln(U_t) + ln(\pi) \qquad (3.9)
\end{aligned}
$$

The estimation of the investment policy function provides the advantage of a control function approach, which is to control for the endogeneity in the technology of skills incorporating a consistent estimate of the heterogeneity component, π. So now, in 3.16 the shocks in period t, η_t, incorporate the consistent estimate of π. For the estimation process, I perform sensitivity analysis using different combinations of the possible exclusion restrictions. The result of this analysis is shown in 3.4.

3.2.3 Unscented Kalman Filter

For the estimation process, I use Mixture of Normals Unscented Kalman Filter algorithm proposed by Cunha et al., 2010. The advantage of the Mixture of Normals Unscented Kalman Filter algorithm is that both the prediction and steps can be accurately approximated by the density of a mixture of normals. To explain, let's use our measurement equation and the technology of skills, but more clearly:

$$
\begin{aligned}
y_t &= g(h_t) + \varepsilon_t \qquad (3.10) \\
h_{t+1} &= f(h_t) + \eta_t, \qquad (3.11)
\end{aligned}
$$

Where $\varepsilon_t \sim N(0, X_t)$ and X_t is the variance-covariance matrix of ε_t, $\eta_t \sim N(0, Y_t)$ and Y_t is the variance-covariance matrix of η_t and $h_t \sim N(m_1, V_1)$ which are the initial conditions for the non-linear filtering problem. Function g from (3.9) includes the factor loadings and the function f from (3.10) the factor coefficients for the evolution of the skill h_t over time.

This procedure aims to estimate the factor loadings, the factor coefficients, the variance-covariance matrix of ε_t and η_t by maximising the conditional likelihood of the model (3.9)-(3.10). Let $p(y)$ denote the conditional likelihood:

$$p(y) = \prod_{t=2}^{T} p(y_t|y^{t-1}), \tag{3.12}$$

For calculating the $p(y_{t+1}|y^t)$ in (3.12) I use nonlinear filtering to obtain a recursive algorithm. For this, I compute the prediction step first, that generates $p(h_t|y^{t-1})$ given that the information about $p(h_{t-1}|y^{t-1})$ applying Chapman-Kolmogorov equation:

$$p(h_t|y^{t-1}) = \int p(h_t|h^{t-1})p(h_{t-1}|y^{t-1})dh_{t-1} \tag{3.13}$$

Then it is necessary to compute the update step using the Bayes's rule:

$$p(h_t|y^t) = \frac{p(y_t|h_t)p(h_t|y^{t-1})}{p(y_t|y^{t-1})} \tag{3.14}$$

and then proceed recursively. Accurately approximation of the $p(h_{t+1}|y_t)$ and $p(h_t|y_t)$ (prediction and update) can be done by the density of a mixture of normals. For the technologies estimation in this chapter I assume a mixture of two normals which writes the density of h_t as[4]:

$$p(h_t|y^t) \cong \tau_t \phi(h_t; m_t, V_t) + (1 - \tau_t)\phi(h_t; m_t, V_t) \tag{3.15}$$

so then using the predicted measured by the mixture $\hat{y}_t = E[g(h_y|y^t)]$ I can compute the right-hand side of (3.12) which is equal to the conditional likelihood.

[4]The calculation of the mean and the variance of the density involve the computation of l integrals, for this I use monomial rules that approximates the values of the mean and variance

3.3 Data and Empirical Specifications

3.3.1 Data

I use new data from Chile that allows researchers to assess the impact of early childhood policies and to provide valuable information for the evaluation and design of social policies in this field since this data contains variables determining children's cognitive and non-cognitive abilities. The *Encuesta Longitudinal de Primera Infancia* (Early Childhood Longitudinal Survey (ELPI)) survey, which is a longitudinal study based on a representative sample of 15,000 children under 5 years old and their families for the first wave, during the second wave the sample is 18,000 as included a refreshing sample from 0 to 3 years old. The first wave was conducted in 2010 and the second one in 2012. It is a representative survey of children from urban and rural areas who were born on January 1, 2006, and August 31, 2009, in the first wave, and that is a representative sample. The sample includes different cohorts of children, distinguished by year of birth. The second wave follows the first wave but also add a refreshing sample of around 3,000 children born on September 1, 2009, and December 31, 2011.

This survey provides demographic information, measures of a set of cognitive, non-cognitive and anthropometrics measures for both the children and the mothers or primary caregivers. There is also information about home assessment using the Home Observation for Measurement of the Environment (HOME) inventory and the Family Care Indicators (FCI) that help to measure intermediate outcomes and mediators for early childhood development. Tests applied in both waves measured the development of children in different areas, such as motor, cognitive, language, emotional and social areas. The following describes the tests used in this chapter as measures of the children's cognitive and non-cognitive skills, maternal cognitive and non-cognitive skills and parental investment for the different technologies specifications. The descriptive statistics are presented in Appendix 5 in Tables A.19 to A.22 for each age stage. In general, the measurements used for t are the ones applied in the ELPI 2010 and for $t+1$ the ones applied in the ELPI 2012.

Children's Cognitive Measures Independently of the child's age the ELPI has information about cognitive skills. For children aged 0-23 months old there is the test *Escala de Evaluación de Desarrollo Psicomotor* (EEDP) which is a Psychomotor Development Rating Scale. The test measures the performance and the reaction of the child to certain situations to be resolved for which a certain level of psychomotor development is required (Rodríguez et al., 2008). The measure used is called fc_eedp in Table A.19. Another test used for the same age range is the *Battelle Developmental Inventory* (BDI 2010). The BDI provides three developmental domains related with cognition: cognition (fc_bdi_cog), communication (fc_bdi_com) and motor (fc_bdi_m). These three domains are further divided into twenty-two separate sub-domains. The motor domain is composed of muscle control, body coordination, locomotion, fine muscle, and perceptual motor. The communication domain includes: receptive and expressive. Finally, the cognitive domain is composed of: perceptual discrimination, memory, reasoning/academic skills, and conceptual development (De la Cruz and González, 1998). For children aged 24-60 months old, the test *Test de Desarrollo Psicomotor (TEPSI)* measures the psychomotor Development Test. TEPSI is a screening test which yields results at global as well as sub-scale levels regarding coordination (fc_tepsi_coo), language (fc_tepsi_l) and motor (fc_tepsi_m) functions (Haeussier and Marchant, 2003 and Wechsler, 1974). The subtest coordination evaluates the ability of the child to take or manipulate objects and draw, through behaviours such as build towers with cubes, threading a needle, recognise and copy geometric figures, draw a human figure among others. The subtest language evaluates aspects of understanding and expression of this, through behaviours such as naming objects, defining words, verbalise or describe actions scenes depicted in films and the subtest motricity evaluates the child's ability to manage their own bodies through behaviours like picking up a ball, hopping, walking on tiptoe or stand on one foot for a while. The previous three tests were applied only in the first wave. The *Test de Vocabulario en Imágenes Peabody, Hispanic America adaptation* (TVIP) is a language skill which is the Spanish version of the Peabody Picture Vocabulary Test (PPVT). This test measures the child's comprehension and understanding of vocabulary using relating words to an illustration since age 30 months old. The scale (fc_tvip_l) should be viewed primarily as an achievement test since it shows the extent of Spanish vocabulary acquisition of the subject. Also, it may be viewed as a screening test of scholastic aptitude (verbal ability or verbal intelligence), or as one element in a comprehensive test battery of cognitive processes when Spanish is the language of the home and community into which the subject was born, has grown up, and resides; and when Spanish is, and has been, the primary language of instructions at school (Dunn et al., 1986). Was applied in both waves. During the second wave, two more tests were applied, the *Battelle Developmental Inventory, Screening Test, 2nd ed.* (BDI 2012) which includes the same areas as the BDI 2010. Has 96 items (two per each age level) extracted from the full version of the BDI and it is a screening test that evaluates the child development from 0-8 years old. The objective is to evaluate the fundamental skills development in three areas in cognition: cognition (fc_bdi_cog), communication (fc_bdi_com) and motor (fc_bdi_m). The second one is the *Test of Learning and Child Development* (TADI) which is a Chilean instrument that allows measuring what children know, and what they do, according to three dimensions of cognition development: language (fc_tadi_l), cognition (fc_tadi_c) and motor (fc_tadi_m), each of which constitutes a separate scale (CEDEP, CIAE, 2012).

Children's Non-cognitive Measures For children aged 0-23 months old there are 2 subscales provided by the *Battelle Developmental Inventory* (BDI 2010) which are personal-social (fn_bdi_ps) and adaptive (fn_bdi_ad). The personal-social domain is composed of: adult interaction, expressions, feelings, affect, self-concept, peer interaction, coping, and social role. The adaptive domain includes attention, eating, dressing, personal responsibilities, and toileting. There is also a scale provided by the *Ages and Stages Questionnaires for socioemotional areas (ASQ: SE)* applied for children 6, 12 and 18 months old. It is a parental report and helps to identify possible problems in the social and affective development of the child. The tests (fn_asq) addresses seven behavioural areas: self-regulation, compliance, communication, adaptive functioning, autonomy, affect, and interaction with people (Squires et al., 2006). The previous socioemotional tests were applied only in the first wave. The *Child Behaviour Checklist* (CBCL1) is applied for children older than 18 months old and obtain details of children's functioning, as seen by parents/caregivers providing results for three general scales: Total Problems, Internalising and Outsourcing There are 7 syndrome scales designates as Emotionally Reactive (fn_cbcl_r), Anxious/Depressed (fn_cbcl_a), Somatic Complaints (fn_cbcl_q), Withdrawn (fn_cbcl_e), Sleep Problems (fn_cbcl_ps), Attention Problems (fn_cbcl_pa), and Aggressive Behaviour (fn_cbcl_ca) (Achenbach and Rescorla, 2000). Was applied in both waves.

Mother's Cognitive Measures The *Wechsler Adults Intelligence Scale* (WAIS) test measures the intelligence/cultural level and memory skills/processing speed/short-term auditory memory through the Vocabulary and Digit Span sub-scales respectively. The WAIS measures human intelligence reflected in both verbal (which measures the subject's knowledge of word meaning) for which we have the measure (pc_wais_voc) and digital distance (ability to recall digits from memory, performance based on the maximum length of a list of digits the subject can recall) abilities for which we have two measures (pc_wais_ds1) and (pc_wais_ds2) (Apfelbeck and Hermosilla, 2000). I also use the mother's years of education of a proxy of cognition.

Mother's Non-cognitive Measures The *The Big Five Inventory* is a self-report inventory designed to measure the Big Five dimensions. The 5 factors are Openness (pn_bfi_o), Conscientiousness (pn_bfi_c), Extroversion (pn_bfi_e), Agreeableness (pn_bfi_a), and Neuroticism (pn_bfi_n). It consists of short phrases with relatively accessible vocabulary (Casullo, 2000).

All Cognitive and Social-Emotional measures have a raw score as well as T scores. As the analysis is within the sample I have internally standardised each measure, taking advantage of the sample size per month, using non-parametric estimation for age and hence I remove the age effect. For this, the first step is to use kernel-weighted local polynomial smoothing methods to regress the raw score on child's age (in months) and recover the mean. The second step is to use again the same estimation method to regress the square of the residuals in the first regression on child's age and recover the variance. Finally, for calculating the z-score, I subtract the mean and divide by the standard deviation for each raw score.

Productivity For initial conditions used in the productivity of the technology of skills I use the anthropometric measures at birth, height and weight, for children that were converted to z-scores using the World Health Organization (WHO, 2006). I also include if the child is male and birth order.

Investment Measurements The Home Observation for Measurement of the Environment (HOME) inventory (Bustos et al., 2001) measures the quality of stimulation and support given to a child in a family atmosphere. The inventory was created by Caldwell et al. (1984, 2003). It consists of 55 items grouped into eight subscales, which records the presence or absence of the trait. This score is obtained from a combination of observation, and the semi-structured interview is conducted in the child's home with the presence of the mother and child. The inventory has eight subscales: Learning Materials, Stimulation of language, Physical environment, Responses from parent to child, Academic stimulation, Modelling and stimulation of social maturity, Diversity of experiences and Acceptance of the child. Nevertheless, in this survey was only collected some items that could be observable during the interview generating new subscales that I will review in the next subsection. Finally, were added three items to observe eating behaviours. The Family Care Indicators (FCI) (Hamadani et al., 2010) is a survey-based indicator of the quality of stimulation of the home environment that help to measure intermediate outcomes and mediators for early childhood development. The FCI was developed to measure home stimulation in large populations and were derived from the Home Observations for Measurement of the Environment (HOME). It has items that can be related to i) parent-child interactions as reading books, singing songs, taking the child outside the home, playing with the child, spending time with the child, etc. and ii) learning materials.

Exclusion restrictions For this I use cross-sectional data from a representative survey *Caracterización Socioeconómica Nacional* (CASEN) to get average wages and unemployment rates by comuna for males and females. The average wages and unemployment rates exploiting variability by comuna. I also use a CPI index per product yearly for monetary investment price by region. The weighted index with information about learning materials such as musical instruments, toys, books, pencils, dolls, educational games, etc.

3.3.2 Empirical Specifications

Specification 1 incorporates only one input for measuring investment, parental investment i_t in child skills at age t, that accounts for several measures, in particular, parent-child activities, learning materials, emotional and verbal responsivity, paternal involvement, acceptance, child's discipline and activities related with the program *Chile Grows with You.* Hence, this specification includes a single-dimensional input for parental investment. $h_{c,t}$ denotes cognitive skills of the child at age t, $h_{n,t}$ denotes non-cognitive skills of the child at age t. P_c and P_n represent maternal cognitive and non-cognitive skills respectively, A is the total factor productivity which is a function of the covariates mentioned in (3.2) and η_t are shocks and/or unmeasurable inputs like the use of childcare or preschool. The technology for skill $k = c, nc$ at period t (where $t = 24 - 47$ months) is:

$$\begin{aligned} h_{k,t+1} &= A_k e^{\eta_t} [\gamma_{k,c} h_{c,t}^{\phi_{c,k}} + \gamma_{k,nc} h_{nc,t}^{\phi_{nc,k}} + \gamma_{k,pc} P_c^{\phi_{pc,k}} \\ &+ \gamma_{k,pnc} P_{nc}^{\phi_{pnc,k}} + \gamma_{k,i} i_t^{\phi_{i,k}}]^{\frac{1}{\phi_k}}, \end{aligned} \tag{3.16}$$

where $\gamma_{k,m} \in [0,1]$, $\sum_m \gamma_{k,m} = 1$ for $m \in [c, nc, pc, pn, i]$. η_t is assumed to be normally distributed and serially independent over all t. This formulation assumes that the formation of skills depend on initial conditions, A_k, the stock of skills in period t, parental investment at t, i_t, maternal skills, P_k, and shocks in period t, η_t.

This specification is estimated with and without using the control function approach described in Section 3.2.2.

Specification 2 is the same as Specification 1, but now it is estimated for different child stages:

- 7-23 months in ELPI 2010; 33-51 months in ELPI 2012
- 24-47 months in ELPI 2010; 44-79 months in ELPI 2012

- 48-58 months in ELPI 2010; 69-83 months in ELPI 2012

In particular, the technology for skill $h_{k,t+1,s}$ for $k = c, nc$ at period t and stage s is:

$$
\begin{aligned}
h_{k,t+1,s} &= A_{k,s}e^{\eta_{t,s}}[\gamma_{k,c,s}h_{c,t,s}^{\phi_{c,k,s}} + \gamma_{k,nc,s}h_{nc,t,s}^{\phi_{nc,k,s}} + \gamma_{k,pc,s}P_c^{\phi_{pc,k,s}} \\
&+ \gamma_{k,pnc,s}P_{nc}^{\phi_{pnc,k,s}} + \gamma_{k,i,s}i_{t,s}^{\phi_{i,k,s}}]^{\frac{1}{\phi_{k,s}}},
\end{aligned} \tag{3.17}
$$

In this way, it is possible to make a distinction between periods of parental investments an the children's stages of development.

Specification 3 incorporates multiple inputs for measuring investment based on material resources and quality time in child skills at age t. The items used for measuring parental investment in material resources are from the FCI, in particular, items that ask about if the child has a space in which to keep their toys and belongings (SpecialPlace), has at least one toy with wheels that can be raised (WheelsToys), there are age-appropriate learning equipment (LearEquip) and has 3 or more books of his/her property (Books3). For measuring quality time with the child these are also using the FCI, in particular, items ask about reading stories or looking at picture books with child (LookBooks), telling stories to child (Stories), taking the child to parks, zoo or museums (GoOut) and spending time with child talking and/or drawing (Drawing). $h_{c,t}$ denotes cognitive skills of the child at age t, $h_{n,t}$ denotes non-cognitive skills of the child at age t. P_c and P_n represent maternal cognitive and non-cognitive skills respectively, A is the total factor productivity and η_t are shocks and/or unmeasurable inputs. The technology for skill $k = c, nc$ at period t (where $t = 24 - 47$ months) is:

$$
\begin{aligned}
h_{k,t+1} &= A_k e^{\eta_t}[\gamma_{k,c}h_{c,t}^{\phi_{c,k}} + \gamma_{k,nc}h_{nc,t}^{\phi_{nc,k}} + \gamma_{k,pc}P_c^{\phi_{pc,k}} \\
&+ \gamma_{k,pnc}P_{nc}^{\phi_{pnc,k}} + \gamma_{k,qt}qt_t^{\phi_{qt,k}} \\
&+ \gamma_{k,i}mr_t^{\phi_{mr,k}}]^{\frac{1}{\phi_k}},
\end{aligned} \tag{3.18}
$$

where qt_t is the quality time the mother and/or father spends with the child at age t and mr_t is material resources in child skills at age t. Table A.21 presents the descriptive statistics for all the investments items available in the survey.

Specification 4 incorporates multiple inputs for measuring investment based on cognitive stimulation and emotional support in child skills at age t. The items used for measuring cognitive stimulation are from the FCI, in particular, items that ask about if the child has toys to push and pull toys (PushPull), has toys for role-playing (RolePlay), there are age-appropriate learning equipment (LearEquip), and there are literary and musical material (Musical). The items for measuring emotional support are from the HOME - Emotional and verbal responsivity scale, in particular, items that ask about if the mother spontaneously vocalises to child at least twice (Vocalises) during the interview, keeps within his visual range and look often (Visual), spontaneously praises child at least twice (Praises) and caresses or kisses child at least once (Caresses). $h_{c,t}$ denotes cognitive skills of the child at age t, $h_{n,t}$ denotes non-cognitive skills of the child at age t. P_c and P_n represent maternal cognitive and non-cognitive skills respectively, A is the total factor productivity and η_t are shocks and/or unmeasurable inputs. The technology for skill $k = c, nc$ at period t (where $t = 24 - 47$ months) is:

$$\begin{aligned} h_{k,t+1} &= A_k e^{\eta_t} [\gamma_{k,c} h_{c,t}^{\phi_{c,k}} + \gamma_{k,nc} h_{nc,t}^{\phi_{nc,k}} + \gamma_{k,pc} P_c^{\phi_{pc,k}} \\ &+ \gamma_{k,pnc} P_{nc}^{\phi_{pnc,k}} + \gamma_{k,ic} cs_t^{\phi_{cs,k}} \\ &+ \gamma_{k,ie} es_t^{\phi_{es,k}}]^{\frac{1}{\phi_k}}, \end{aligned} \tag{3.19}$$

where cs_t is the cognitive stimulation provided by parents for the child at age t and es_t is emotional support provided by parents in child skills at age t. Table A.21 presents the descriptive statistics for all the investments items available in the survey.

3.4 Results

This section presents the results of the multiple analysis used for the estimation of the different specifications of the technology of skill formation. The descriptive statistics of the data used for the specifications are presented in Appendix 5 in Tables A.19 to A.22.

3.4.1 Measurement error: Noise and Signal

Before going throughout the results of the several specifications for the technology of skill formation estimation in early childhood I analyse the importance of the measurement error. Section 3.2.1 present the measurement equations from which is possible to calculate the variance of $y_{k,t,j}$ for each k, j at t:

$$var(y_{k,t,j}) = \alpha^2_{k,t,j} var(ln(h_{k,t})) + var(\varepsilon_{k,t,j}),$$

which can be decomposed by the signal $\alpha^2_{k,t,j} var(ln(h_{k,t}))$ and the noise $var(\varepsilon_{k,t,j})$. To calculate the fraction of the variance of the measure $y_{k,t,j}$ due to signal, $s_{k,t,j}$, and noise, $n_{k,t,j}$, we just need to divide by the $var(y_{k,t,j})$ both the signal and noise:

$$\begin{aligned} s_{k,t,j} &= \frac{\alpha^2_{k,t,j} var(ln(h_{k,t}))}{var(y_{k,t,j})} \\ n_{k,t,j} &= \frac{var(\varepsilon_{k,t,j})}{var(y_{k,t,j})}, \end{aligned}$$

Table A.23 in Appendix 5 shows the results of this analysis. For child's cognitive skills at t most of the measures show acceptable (above 75%) percentage of signal, the same occurs for child's cognitive measures at $t+1$, only the measure presents a value lower than 75% which is the measure fc˙tadi˙m which is a proxy of motor skills, the signal is 61.3%. Similar results hold for child's non-cognitive measures at t and $t+1$ which are above 63.5% except one measure that has a signal of 44.0%. Mother's cognitive and non-cognitive measures also show acceptable (above 71%) percentage of signal except for the measure of the Big Five Inventory which measures Agreeableness (pn_bfi_a) with a signal of 52.8%.

The measures for investment present different results: parent-child activities and learning materials are quite good (above 79%) instead emotional and verbal responsivity and paternal involvement have a signal of around 48.0% meanwhile the measures of acceptance, child's discipline and activities related to the program *Chile Grows with You* are really bad with a noise above 90.0%. Figure A.0.2 in Appendix 5 show the factor loadings derived from the estimation process in particular for *Specification 2* once dealing for the endogeneity of inputs.

The items used for measuring parental investment in material resources have a signal above 73.0%. For measuring quality time with the child only two of them present high signal percentage (85%), in particular, the measures that ask about reading stories or looking at picture books with child (LookBooks) and telling stories to child (Stories), instead the measures that ask about taking the child to parks, zoo or museums (GoOut) or spending time with child talking and/or drawing (Drawing) have a signal between 20-30%. The items used for measuring cognitive stimulation have a signal above 78.0%. For measuring emotional support, only one present high noise percentage (68.7%) which measures if the mother caresses or kisses child at least once (Caresses) during the interview, for all the rest the signal is above 61.0%.

3.4.2 Parental Investment Equations and Identification

Given the endogeneity problem that exists in the estimation of the technology of human capital it is crucial to use exclusion restrictions that are correlated with parental investment but which do not suffer from endogeneity and therefore obtain consistent parameter estimates.

Table 3.1 presents the estimates that are derived from the estimation of the parental investment equations using the multiple dimensions for measuring parental investment which are the one-dimensional and two-dimensional inputs. For the latter, I also use two types, first, material resources investment and quality time investment and secondly, cognitive stimulation investment and emotional support investment. The results show that child's skills and maternal's skills are significant and explain important part of the reason why parents invest in their children for all the specifications except for the Emotional Support. These findings explain the existence of skill gaps even at early ages and conditional on initial levels of skills in children as mothers with a higher level of cognitive and non-cognitive skills invest more in their children. I also find that investments decrease in any dimension if the level of unemployment increases. I find that female wages have a positive impact on investments, particularly, on material and emotional support. The effect of prices is not significant and it has a small impact on investments.

I perform sensitivity analysis using different combinations of the possible exclusion restrictions and I find that all of them work as instruments. Nevertheless, there is one parental investment equation, the one for emotional support, which do not present instruments that are too strong as the ones for the rest of investment equations.

Table 3.1: Parental Investment Equations, Age 24-47 mths

	Log Investment	Log Material Resources	Log Quality Time	Log Cognitive Stimulation	Log Emotional Support
Cognitive Skill $\gamma_{t,c}$	0.246	0.142	0.102	0.114	0.019
	(0.017)	(0.019)	(0.018)	(0.018)	(0.019)
Non-cognitive Skill $\gamma_{t,nc}$	0.047	0.059	0.026	0.043	0.110
	(0.015)	(0.017)	(0.016)	(0.016)	(0.016)
Mother's Cognitive Skill γ_{pc}	0.159	0.140	0.049	0.149	0.019
	(0.014)	(0.016)	(0.015)	(0.015)	(0.016)
Mother's Non-cognitive Skill γ_{pnc}	0.124	0.179	0.150	0.158	0.105
	(0.016)	(0.018)	(0.017)	(0.017)	(0.018)
Log unemployment U_t	-0.102	-0.109	-0.044	-0.106	-0.002
	(0.015)	(0.017)	(0.016)	(0.017)	(0.017)
Log average female wages $F_{W\,t}$	0.070	0.112	-0.027	0.086	0.125
	(0.024)	(0.027)	(0.018)	(0.026)	(0.027)
Log average male wages $M_{W\,t}$	0.016	0.009	0.065	0.047	-0.049
	(0.024)	(0.027)	(0.026)	(0.026)	(0.027)
Log investment prices P_t	0.001	0.032	-0.011	0.033	-0.012
	(0.015)	(0.017)	(0.016)	(0.016)	(0.017)
Unobserved Heterogeneity π	-1.870	0.896	1.290	-2.389	-2.521
	(1.092)	(0.124)	(1.067)	(0.730)	(0.826)
F-test (p-values) $F_{W\,t}$ and U_t	27.97	30.52	4.06	27.54	11.23
	0.0000	0.0000	0.0172		0.0000
F-test (p-values) $F_{W\,t}$, $M_{W\,t}$, U_t and P_t	24.30	30.28	5.27	33.96	9.55
	0.0000	0.0000	0.0004	0.0000	0.0000

Note: Standard errors in parentheses are obtained through bootstrapping.

3.4.3 Estimation of the Technology of Human Capital Formation: Specifications

For the estimation of the parameters of the technology of human capital for different stages in childhood, I use the estimation strategy outlined in Section 3.2.3. In all estimations, standard errors are obtained through bootstrapping. Table 3.2 show the estimates of the cognitive and non-cognitive skill CES technologies for children at t aged between 24-47 months and between 44-79 months at $t+1$ based in *Specification 1*, the first two columns present the result for cognitive first without using the control function approach for dealing with the endogeneity of investment meanwhile the second use this method. The same follows for columns 3 and 4. The comparison of the estimations with and without control function approach shows that the effect of self-productivity decreases for the formation of child's non-cognitive and instead the cross-productivity increases (column 3 against 4).

Table 3.2: Production Function of Cognitive and Non-cognitive skills: One investment input, Age 24-47 mths, Specification 1

	Cognitive *t+1*		Non-Cognitive *t+1*	
	without control fn	with control fn	without control fn	with control fn
Cognitive Skill $\gamma_{t,c}$	0.605	0.692	0.123	0.339
	(0.027)	(0.059)	(0.070)	(0.062)
Non-cognitive Skill $\gamma_{t,nc}$	0.038	0.087	0.405	0.182
	(0.012)	(0.040)	(0.023)	(0.055)
Mother's Cognitive Skill γ_{pc}	0.034	0.016	0.019	0.003
	(0.043)	(0.008)	(0.011)	(0.005)
Mother's Non-cognitive Skill γ_{pnc}	0.007	0.005	0.165	0.244
	(0.004)	(0.030)	(0.067)	(0.010)
Investment γ_i	0.200	0.317	0.231	0.288
	(0.028)	(0.075)	(0.040)	(0.034)
Complementary parameter ϕ	0.106	0.543	0.012	0.788
	(0.121)	(0.113)	(0.054)	(0.148)
Elasticity of substitution $1/(1-\phi)$	1.119	2.187	1.012	4.715

Note: Standard errors in parentheses are obtained through bootstrapping.

For the formation child's cognitive skills self and cross-productivity increase (column 1 and 2), mother's cognitive skills decrease in the formation of both child's skills meanwhile mother's non-cognitive skills increases in the formation of child's non-cognitive skills. There is substantial evidence of the effect of parental investment in early childhood development and also support the fact that parental investment is endogenous.

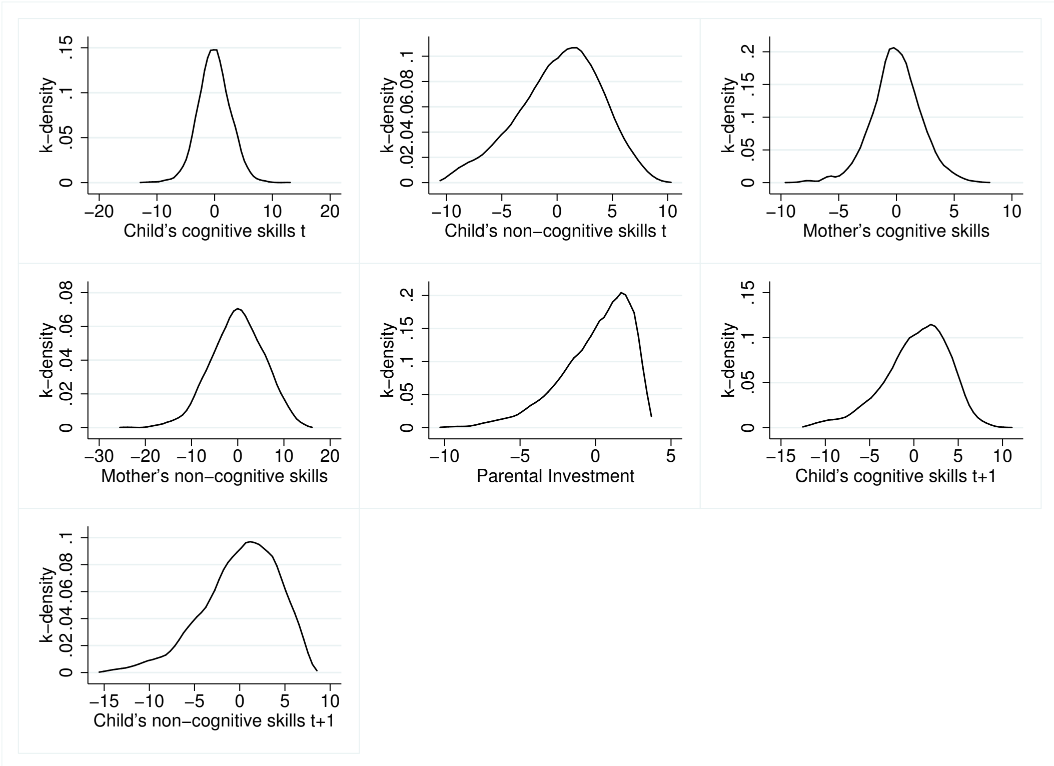


Figure 3.1: Kernel densities of latent traits: One investment input, Age 24-47 mths, Specification 1 with control function

Comparing now the formation of cognitive and non-cognitive skills in children under the control function approach, the results provide evidence about the importance of the stock of child's skills as well as early investment in childhood development. Regarding self-productivity, there is evidence for the formation of both child's cognitive and non-cognitive skills, but the effect is stronger for cognitive skills, which means that an increase in the stock of skills at t have more prominent effects on the formation of the same skills at $t + 1$. Mother's skills have relative low persistence in future cognitive skills, instead of, for future child's non-cognitive skills the mother's non-cognitive skills have a more significant effect. The complementary parameter is bigger for future skills once controlling for endogeneity and the elasticity parameter is near to 1 for both skills without controlling for endogeneity which implies a Cobb Douglas form instead once using the control function approach we reject that the technology has a Cobb Douglas form.

Figure 3.1 show the densities of the latent traits derived from the measurement error system and used for the estimation of the cognitive and non-cognitive skill CES technologies for children at t.

Specification 2 is the same as Specification 1 controlling for endogeneity, but now it is estimated for different child stages at t. Table 3.3 show the estimates of the cognitive and non-cognitive skill CES technologies for children at t with 7-23 months in the first column, with 24-47 months in the second column and with 48-58 months in the last column for each child's skill. In terms of the formation of child's cognitive skills, the results are similar to the ones presented previously for children aged between 24-47 months but the more significant difference is how parental investment foster cognitive skills between 24-47 months with respect to early and older stages instead for future non-cognitive skills the parental investment have the same effect for all the age stages. There is evidence of cross-productivity for both skills which raises for the later stages.

Table 3.3: Production Function of Cognitive and Non-cognitive skills: One investment input, different age stages, Specification 2

	Cognitive *t+1*			Non-Cognitive *t+1*		
	7-23 mths	24-47 mths	48-58 mths	7-23 mths	24-47 mths	48-58 mths
Cognitive Skill $\gamma_{t,c}$	0.799	0.692	0.777	0.280	0.339	0.336
	(0.039)	(0.059)	(0.025)	(0.047)	(0.062)	(0.034)
Non-cognitive Skill $\gamma_{t,nc}$	0.099	0.087	0.167	0.233	0.182	0.198
	(0.028)	(0.040)	(0.035)	(0.059)	(0.055)	(0.070)
Mother's Cognitive Skill γ_{pc}	0.054	0.016	0.028	0.003	0.003	0.004
	(0.020)	(0.008)	(0.010)	(0.007)	(0.005)	(0.08)
Mother's Non-cognitive Skill γ_{pnc}	0.009	0.005	0.011	0.274	0.244	0.239
	(0.025)	(0.030)	(0.021)	(0.031)	(0.010)	(0.032)
Investment γ_i	0.039	0.317	0.017	0.210	0.288	0.223
	(0.002)	(0.075)	(0.018)	(0.046)	(0.034)	(0.068)
Complementary parameter ϕ	0.522	0.543	0.655	0.875	0.788	0.839
	(0.177)	(0.113)	(0.157)	(0.228)	(0.148)	(0.177)
Elasticity of substitution $1/(1-\phi)$	2.094	2.187	2.896	8.004	4.715	6.229

Note: Standard errors in parentheses are obtained through bootstrapping.

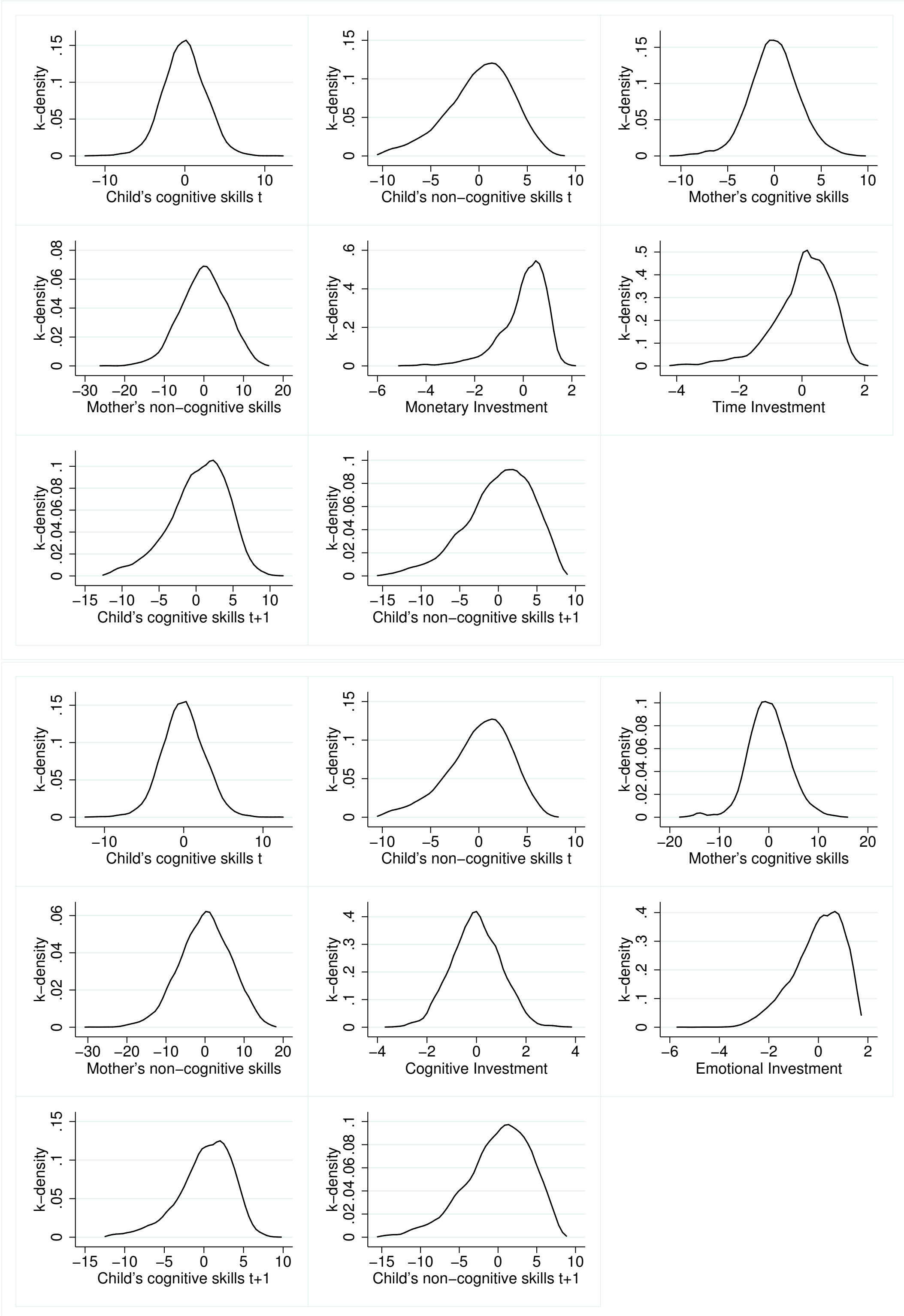


Figure 3.2: Kernel densities of latent traits: Multiple investment, 24-47mths, Specifications 3 and 4

Table 3.4: Production Function of Cognitive and Non-cognitive skills: Multiple investments, Age 24-47 mths, Specification 2, 3 and 4

	Cognitive *t+1*			Non-Cognitive *t+1*		
	Inv	Money/Time	Cog/Non-cog	Inv	Money/Time	Cog/Non-cog
Cognitive Skill $\gamma_{t,c}$	0.692	0.664	0.739	0.339	0.269	0.214
	(0.059)	(0.052)	(0.042)	(0.062)	(0.058)	(0.044)
Non-cognitive Skill $\gamma_{t,nc}$	0.087	0.075	0.074	0.182	0.108	0.091
	(0.040)	(0.027)	(0.033)	(0.055)	(0.051)	(0.043)
Mother's Cognitive Skill γ_{pc}	0.016	0.021	0.012	0.003	0.003	0.002
	(0.038)	(0.010)	(0.007)	(0.005)	(0.005)	(0.002)
Mother's Non-cognitive Skill γ_{pnc}	0.005	0.004	0.003	0.244	0.037	0.000
	(0.030)	(0.012)	(0.015)	(0.010)	(0.012)	(0.018)
Investment $\gamma_{t,m}$	0.317			0.288		
	(0.075)			(0.034)		
Material resources Investment γ_{mr}		0.123			0.135	
		(0.032)			(0.046)	
Quality time Investment γ_{qt}		0.112			0.223	
		(0.028)			(0.031)	
Cognitive Stimulation Investment γ_{cs}			0.091			0.097
			(0.022)			(0.031)
Emotional Support Investment γ_{es}			0.081			0.235
			(0.042)			(0.098)
Complementary parameter ϕ	0.543	0.448	0.596	0.788	0.778	0.787
	(0.113)	(0.135)	(0.124)	(0.148)	(0.236)	(0.224)
Elasticity of substitution $1/(1-\phi)$	2.187	1.811	2.477	4.715	4.499	4.706

Note: Standard errors in parentheses are obtained through bootstrapping.

Specification 3 and *Specification 4* incorporate multiple input for measuring investment. Table 3.4 show the estimates of the cognitive and non-cognitive skill CES technologies for children at t with 7-23 months for *Specifications* 1, 3 and 4 in column 1, 2 and 3 respectively for both future child's skills. There is still evidence of self-productivity for future child's skills, though the effect of self-productivity is bigger for non-cognitive skills. There is also evidence of cross-productivity for the future non-cognitive skills. As before, mother's cognitive skill has low persistence in both child's future skills and the mother's non-cognitive skills affect only for fostering the non-cognitive skills, but this effect disappears once that the emotional support investment is incorporated in the estimation. In terms of the effects of separating the investment in material resources and quality time in child skills at age t the results show that material resources is important for determining future child's cognitive skills and quality time for determining future child's non-cognitive skills which are the similar result found by Attanasio et al. (2015).

Figure 3.2 shows the densities of the latent traits derived from the measurement error system and used for the estimation of the cognitive and non-cognitive skill CES technologies for children at t for *Specifications* 3 and 4. Regarding the effects of separating the investment in cognitive stimulation and emotional support in child skills at age t, the results show that there is not much return regarding cognitive stimulation meanwhile the return of emotional support is higher on future child's non-cognitive skills.

3.5 Conclusions and further work

During the last years, an increasing body of research has focused on the development of early human capital mostly because gaps in early skills translate into long-term gaps in social and economic inequality. Unfortunately, the human capital formation is a complicated process which is a multi-dimensional and dynamic process, the dimensions of human capital interact both within and across periods, and several unobserved inputs are crucial for estimating correctly the technology of human capital. Some of the key questions are how the skills develop in children's human capital? What is "true" technology of skill formation and how does it change? Are the inputs of the technology malleable? Do public policies influence child's outcomes through the inputs of the technology of skill formation?

This chapter provides some evidence for answering the previous answers. Exploiting the rich panel structure of the *Encuesta Longitudinal de Primera Infancia* (Early Childhood Longitudinal Survey (ELPI)) survey I find evidence about the importance of the stock of child's skills as well as early investment in childhood development. Comparing the formation of cognitive and non-cognitive skills in children dealing or not with endogeneity, there is substantial evidence of the effect of parental investment in early childhood development and also support the fact that parental investment is endogenous. Based on the estimation of the same production function but for different age stages, the primary result is how parental investment foster cognitive skills between 24-47 months concerning the first and the latest stage instead for future non-cognitive skills the parental investment have the same effect for all the age stages. There is evidence of cross-productivity for both skills which raises for the latest stage. Regarding the effects of separating the investment in material resources and quality time in child skills at age t, the results show that material resources are essential for determining future child's cognitive skills and quality time for determining future child's non-cognitive skills. Finally, splitting the investment in cognitive stimulation and emotional support in child skills at age t, the results show that there is not much return regarding cognitive stimulation meanwhile the return of emotional support is higher on future child's non-cognitive skills.

Nevertheless, there are still some limitations in the analysis and further work can be done, for example, using the structural parameters of the technology for simulating policies as well as attempt to characterise the process of human capital accumulation in early years for multiple dimensions (adding health or language skills) to analyse if the return of investment is biased or not.

Chapter 4

Parental Beliefs and Investments in Human Capital

4.1 Introduction

The importance of what happens in the first years of life for human development has been extensively documented. When children grow up in poverty, both in developed or developing countries, children are exposed to a variety of risk factors that it is likely to contribute the accumulation of considerable lags in many dimensions, including cognitive and socio-emotional skills. These lags and delays have substantial long-term consequences for human development and well-being.

One of the factors that seem to be extremely important in affecting children development is the level of parental investment, which can take the form of small amount of time parents spend in quality interactions with their children or low levels of didactic material investments. [Attanasio et al. (2015)] find that parents invest "too little" using an experiment in a developing country which is consistent with anthropological and sociological studies in the US ([Lareau (2003)] and [Putnam (2015)]).

A potentially important question, therefore, is what drives parental investment. While financial and material resources are undoubtedly essential and a key determinant, they might not be the only or not even the most crucial factor that determines low levels of parental investment in low-income families. This consideration is particularly true if one thinks of the quality time parents spend with their children, where quality is defined regarding interactions and stimulating activities parents do with their children.

An intriguing possibility is that parents choose low levels of parental investment because, rightly or wrongly, they perceive the returns to that investment as being low. In this case, while all parents might care equally about the development and well-being of their children, low-income parents might not be aware of the importance that some specific activities, such as talking and interacting in specific ways with a small child, might have for their development. And yet, the findings in child development indicate that early stimulation is essential for subsequent development and that exposure to language and meaningful interactions drive subsequent developments. Under this hypothesis, low-income parents, pursuing what [Lareau (2003)] defines 'natural growth' could be making the "wrong" choices.

The standard practice in economics to answer what drives parental investment is to estimate (dynamic) models where parents are assumed to "know" the production function of child development or human capital. Within such models, parental investment is driven by the nature of the production function of human capital, by financial resources and the cost of investment in children and by how much parents care for their children. If parents have the 'wrong' idea about the nature of the production function, these models could result in misleading conclusions about the nature of investment and about the design of policies aimed to improve child development.

One possibility to study parental behaviour without assuming that parents know the nature of the process of child development or the production function of human capital is to elicit directly parental beliefs about the process of child development and, in particular, about the usefulness of stimulation and investment and how these inputs interact with initial conditions.

In this paper, we develop some of the ideas in [Cunha et al. (2013)] and elicit parental beliefs in a sample of poor mothers in Colombia. We show how to convert the answers to a specific set of questions into estimates of expected rates of returns on specific investment and then relate these estimates to actual parental investment behaviour. With this approach, under the assumptions that mothers know the shape of the production function but might have the wrong idea about the productivity of specific inputs, we can convert the answers mothers give to specific questions into parameter values for the perceived production function. We also formulate and estimate a model in which parents have objective and subjective beliefs about the technology governing the formulation of skills in early childhood development. We can then use this model to perform a number of counterfactual experiments.

In general, the production function for skills approach in early childhood assumes that parents have rational expectations about the process that determines how their children in their early years develop cognitive and non-cognitive skills. This assumption is very tough because implies that all parents know the objective distribution function of an extra increase in the investment in their children's development, irrespective of parents' background. The state of the art in the estimation of a skill formation process relies on a methodology developed by Cunha et al. (2006, 2010), which account for endogeneity of parental investment incorporating an independent shock over people, but not over time and assuming it will be realised before parents make investment choices, so it is expected parental investment to respond to it. Hence, the variation in observed investments across families is not attributed to the heterogeneity in the maternal knowledge about child development.

The goal of this chapter is to shed light on the importance of maternal subjective expectations in explaining the heterogeneity in the maternal choices of investments in the development of their children. Subjective expectations about the production function of skills in early childhood development are crucial since mothers may have biased expectations about the returns to investments, which is crucial to pin down in designing policies aimed at remediating poor investments. To determinate the importance of maternal subjective expectations, we will formulate and estimate a model with mother's preferences, the budget constraint faced by the mother and a production function in the spirit of Cunha et al. (2006, 2010).

We apply this methodology to data collected in the context of an intervention in Colombia, in which home visitors paid weekly visits to randomly chosen households with the aim of promoting child cognitive and psychosocial development and improving mother-child interactions. The intervention targeted households with children aged 12 to 24 months at baseline and lasted 18 months. Very rich data containing measures of skills and investments were collected both before and after the intervention of the households that received home visits and those that did not. But, if we only observe measures of human capital and investments, it is impossible to decompose heterogeneity in expectations from heterogeneity in preferences (Manski, 2004).

For this reason, it is necessary to collect subjective maternal expectations data about the production function for skills. The method we used to collect this information will help overcome this identification challenge. To do so, there was an adaptation of the Motor and Social Developments Score (MSD) used in the CNLSY/79 survey for the second follow- up of the intervention in Colombia. The main characteristic of the instrument is that mothers provide age answers for different types of child language development (high and low) and different levels of investments (high and low). Hence, mothers faced four different scenarios. In this context, the instrument has three sections; the first one was used as a trial to explain to the mothers how the instrument works and collect age answers for each hypothetical scenario asked. The second section collects maternal expectations about monetary and time investment, for example, the question with low development and low investment asks: if a child of 9 months, that is less developed in its language, can say easy words for this age like "guaugu bye or babe" and he(she) still cannot say difficult words like "dog, shoe or cow". If the mother of this child only spends two hours a day reading and telling stories, talking and playing with him/her and purchase a few educational materials, at what age do you think this child can say 3 easy/middle/hard words such as hand, bed and sleep/ glass, door and play/ shirt, broken and flower? The third section collects maternal expectations about preschool investment.

This methodology has several advantages. The first one is comparability; as the same words used to maternal subjective expectations about child development are the same one that will be used to measure language development in the production function. Secondly, the different scenarios allow us to assess the sensitivity of answers concerning variations in them. Thirdly, we measure subjective expectations as the mother is always asking "what age do you think this child". Finally, as the questions are asked for different difficulty levels (easy/middle/hard), we will be able to infer the mothers' subjective probability.

The contribution of this chapter is twofold. By shedding light on the importance of maternal subjective expectations in explaining the heterogeneity in the maternal choices of investments in the development of their children, our analysis will inform the design of optimal strategies to remedy early deficiencies in skills and investments. Moreover, to the best of our knowledge, this paper will be the first one to exploit information from a randomised control trial to estimate a model of skill formation, incorporating maternal subjective expectations, thus representing a significant contribution to the academic literature on the estimation of skill production functions.

4.2 Parental beliefs about child development

In this section, we describe the methodology we used to elicit information about parental beliefs. Our main data set was collected in the second follow up of a Randomised Controlled Trial performed to improve parental practices in Colombia. The short run impacts of this intervention are described in [Attanasio et al. (2014)], which report a 0.26 sd increase in the cognition of children aged 30 to 42 months at the end of the intervention. The mechanisms behind these impacts are studied in [Attanasio et al. (2015)], who argue that an important part of this impact was driven by a substantial increase in parental investment during the intervention period.

Two years after the end of the intervention, it was decided to collect an additional follow-up to asses the medium run impacts of the intervention on child development. Within this data collection exercise, a module eliciting parental beliefs about the process of child development was added to main survey. We now describe the questions we used to elicit belief and provide some information about the approach used to construct the questionnaires.

The main idea in the construction of the questions about parental beliefs, is that child development, in terms of language and cognition, can be represented by a factor H which evolves as the child ages. As the child reaches different degrees of development, he or she achieves the ability to perform certain tasks, such as the understanding certain words and the ability of using such words.

This developmental process depends, on average, on where the child starts from at a given age and on parental investment. We also assume that the mothers whose beliefs we elicit share the assumption that child development depends on some initial condition and parental investment. If we denote with H_1 child development at the end of the period considered, with H_0 with the initial condition and with X parental investment, we can summarise this assumption as follows:

$$E[H_1|I_i] = E[f^i(H_0, X) + \epsilon_i|I_i] = f^i(H_0, X) \tag{4.1}$$

where the superscript i indicates the mother, the variable ϵ_i heterogeneity in the process of development, with zero mean, and the variable I_i denotes mother i's information set. Notice that we assume that each mother might have a different idea about the process of development that links H_0 and X to H_1.

In order to map the beliefs of each mother, as represented by equation (4.1) we present the mother being interviewed with some questions about a hypothetical child and ask her about the process of development of such child under differ scenarios. The interviewer stressed that the hypothetical child was not the woman's own child but a generic one, as such approach made the description of the hypothetical scenarios easier.

Given the structure in equation (4.1), the scenarios are supposed to describe different values of H_0 and X and the possible answers are framed to present different possible values of H_1. In particular, we design four different scenarios for H_0 and X, corresponding to relatively low and high value of X and H_0. Moreover, mothers are asked about the age at which the child is able to achieve certain tasks under each alternative scenario. As we argue below, this should correspond to different expected values of H_1 for the different scenarios.

In the rest of this section, we discuss how the scenarios are formulated and how they relate to H_0 and X and how the answers given for each scenario relate to H_1. Before that, however, we sketch how the abstract constructs of H_1, H_0 and X are translated into variables and situations that can be easily understood by the mothers we interview.

4.2.1 Latent variables and measurement

The main idea in relating observed variables to the abstract constructs that enter equation (4.1) is that of straightforward latent variable model. We assume that, corresponding to each of the three variables H_1, H_0 and X, we have a number of observable indicators whose behaviour is affected by one of our three variables and some measurement error. For instance, given a variable m^j relating to factor H_0, we have the relation:

$$m^j = g^j(H_0, \epsilon^j) \tag{4.2}$$

The nature of the relation 4.2 depends on the type of variable we observe. For instance, if consider the ability of a child to say or understand a number of certain words, it is natural to use an Item Response Theory model, while other continuous variables will require other latent variables model. It might be also necessary to combine IRT and other models.

Estimates of these latent variable models can then be used to convert a certain set of observations on some or all the variables used to estimate the model, to derive an estimate of the corresponding latent factor. In what follows, we will be precise about the type of estimator of the factor scores used, but at this point we only want to convey the idea behind our approach.

While for X and H_0 the use of a factor model is straightforward, relating H_1 to the answers given by mothers is more complicated, as the answers report the age when a child is expected to perform certain tasks. As we discuss below, we assume that development progresses with a child's age, although at different paces. We therefore we will need to use the definition of developmental age.

4.2.2 Construction of Scenarios

To choose a design for possible scenarios, we use evidence for the families in our sample, as observed in the firs follow up survey. In particular, for H_0 scenario, we use the McArthur Language Inventory - First (MLI- I) test, which report the ability of a child of a certain age to say 100 words. We use these data to estimate a two-parameter Rasch model, with the two parameters measuring the difficulty α^j of word j and the saliency β^j of the same word. We then choose some words for a relative high saliency as being particularly representative of

H_0. We also choose both words that are estimated t be particularly easy and words that are estimated to be hard. We then define two scenarios for the initial condition for the hypothetical child as 'lo' if this child can only say the easy words, while as 'high' if he or she can understand and say both the easy and the difficult words.

Analogously, we construct 'low' and 'high' scenarios for X, which is the latent factor that represents parental investment. The scenario in which X is 'low' is the one in which the mother spends little time with the child and spends little money on didactic materials. The scenario in which X is "high" is the one in which the mother spends more time and money on the child's development.

To identify appropriate variables and value, we use data available in the first follow up survey that measure the Family Care Inventory (FCI) and estimate a latent factor model. We use items measuring 'activities' between the mother and the child, in particular, reading and telling stories, talking and playing with him/her. We also use the time measured in hours describing the amount of time that the mother spends doing these activities with the child. We also use items measuring the provision of 'didactic materials' to the child, in particular, toys that induce physical movement (e.g., ball) and dolls for "low" levels of investment and toys to ensemble or build things, things for drawing, painting and/or writing, children books and toys for learn shapes and/or colours for "high" levels of investment. Analysis of these data through an IRT model allows us to determine what constituted "low" and "high" levels of investment.

The scenarios for investment were presented to the mothers in laminated illustrated cards so that the oral description could be reinforced with visual stimulation. Figure A.0.4 in the Appendix 6 shows the vignettes used during the elicitation of beliefs.

Having constructed the scenarios for X and H_0, we present to mothers four scenarios crossing the 'low' and 'high' values of investment and initial condition. We therefore have a (low, low) scenario, a (low, high) , a (high, low) and a (high, high) scenario.

4.2.3 Survey Questions

For each of these four scenario, the mother was asked at what age a hypothetical child would start saying 3 sets of words in that particular scenario. The three set of words were chosen to be easy (first set), medium (second set) and hard (third set). These words were chosen from the McArthur Language Inventory - Second (MLI-II) after performing IRT analysis on baseline data for children between 19-24 months described in Section 4.5.1.2. All the chosen words had relatively high loading factors' β^j's. Easy words had low intercepts (α^j's), hard words had high α^j's and medium words medium α^j's. For ease of notation, we will label the easy words as having $\{j = e_1, e_2, e_3\}$, the medium words as having $\{j = m_1, m_2, m_3\}$ and the hard words as having $\{j = h_1, h_2, h_3\}$.

The reason to ask about the ability to understand and say different groups of words for each of the four scenarios, is to construct multiple measures of the perceived outcome H_1. As we expect that these answers are affected by measurement error, using different groups of words with different levels of difficulty allows us to use these observations to average out measurement errors, which, for different levels of difficulty are likely to be independent across measures.

To indicate their answers, mothers used wooden tablets that had been marked with different ages at the top and that contained a number of ropes with a bead. For each set of words and each scenario, the mother would put the corresponding bead at a given age. At the end of the exercise each mother was presented with two wooden tablets (corresponding to the four scenarios), each with the six ropes and beads (the first/second tablet for Low/High Initial Condition with 3 ropes for low investment and 3 ropes for high investment) and was asked whether she would want to revise any of the questions. Answers about Maternal Investment are showed in Table 4.1.

Consistency (in that easier words - or more investment- should correspond to earlier ages) was not forced.[1] This procedure was chosen after extensive

[1] As a practice, before the developmental questions, mothers were asked about a younger child and when he/she would start to crawl, walk and run using two different scenarios in terms of nutrition. During these practice questions, the interviewer would point out to inconsistencies, if, for instance, the respondent would indicate that the child would start to run before initial conditions (less or more healthy). The point of this was to familiarise the respondents with the method.

Table 4.1: Answers about Maternal Investment

		Mean	St. Dv.	Min	Max
Low Initial Conditions	Low Investment	18.3	6.2	9	48
		23.5	7.3	10	48
		29.5	8.8	11	48
	High Investment	15.8	5.7	9	48
		20.1	6.8	9	48
		25.0	8.2	9	48
High Initial Conditions	Low Investment	14.4	4.8	9	48
		18.0	5.6	9	48
		22.3	7.2	10	48
	High Investment	13.5	5.3	9	48
		16.7	5.9	9	48
		20.3	7.2	9	48

piloting in which alternative wordings of the questions were tried.

4.3 Using the beliefs questions

In this section, we show how the answers to the beliefs questions can be used. A simple way to use the questions is to compute, for each of the two initial conditions, the perceived return to investment measured in terms of time. The exercise can be done for the three group of words used as an outcome (easy, medium or hard) or taking an average of the three. For instance, a mother might think that for a well-developed hypothetical child an increase of investment going from the low to the high investment will speed the learning of a given set of words by 2 months.

However, if we want to use the answers to the beliefs questions to compare the process of development as perceived by mothers to actual data, we need to rescale the answers received and the scenarios to make them comparable to the actual development and investment measurements available on a group of children.

Our procedure involves 5 steps. The first few steps are relatively simple and do not involve strong assumptions. in particular, the first two steps are performed on actual data and do not involve the use of the beliefs questions.

The first step rescales a widely used index of development in terms of developmental age. This procedure is useful because the beliefs questions are asked in terms of age, so that the scaling and comparison with actual data becomes simpler. The second step consists in estimating a latent factor model with observable variables that makes useful to estimate X, H_0 and H_1 given a (sub) set of measures.

Step 3 uses estimates from the step 2 to convert the answer to the beliefs questions into estimates of H_1. At this point we also obtain prediction of H_0 and X given the four scenarios. In the step, we also use the normalisation obtained in step 1 to express H_1 in terms of developmental age. In step 4, we use the estimates in step 3 to compute the return to investment in terms of developmental age.

As we move forward towards the comparison of the beliefs data with the correlations between development progress, inputs and initial conditions observed in actual data, stronger assumptions are necessary to make the available information comparable. In step 5, for a given functional form assumption on equation (4.1), we obtain, for each mother, an estimate of the parameters of the production function, which can be compared to those obtained with actual data. For this last step to be meaningful the appropriate scaling for the various measures we use is important. In the rest of this section, we describe the 5 steps in detail.

4.3.1 Scaling: The Relationship Between $H_{i,0}$, $H_{i,1}$ and Age

A test that is often used to measure the development of children aged below 42 months is the Bayley Scales of Infant Development (BSID). The BSID is made of three scales, measuring cognition, receptive language and expressive language. Each BSID scale is measured in terms of the raw score based in the number of items for which the child receives credit (i.e., 1 point). The BSID was also collected in our sample and we include it, together with the MLI tests, in the factor model we use to identify the latent factor H. Indeed, we use the BSID to provide a unifying metric. However, given that the answers to the development questions included in the beliefs module were formulated in terms of age, we first convert the BSID scale into what we define as 'developmental age'.

Let $BSID^{raw}_{i,j}$ denote the observed raw score for child i in scale j. Let a_i denote the logarithm of the child's age at the time that child development is measured. For each scale, we calculate the mean of the raw score for all the children in our data of a certain age a_i, $mean(BSID^{raw}_{i,j})$. We then regress log age on an intercept and these average raw scores. We denote the intercept and slope of this regression, which converts average scores into 'age' as α^j and β^j and use them to define the developmental age of each child corresponding to scale j.

$$devage_{i,j} = \alpha^j + \beta^j * ln(BSID^{raw}_{i,j}), \tag{4.3}$$

Notice that equation (4.3) converts the BSID score of every child into an estimated developmental age. This equation can be used to normalise in the same fashion both actual data and the data derived from the beliefs questionnaires.

4.3.2 Using information from a latent factor model

As mentioned above, the beliefs questions ask the age at which a hypothetical child can understand or say three groups of three words, chosen on the basis of the difficulty and saliency parameters of an estimated latent factor model. Such a model includes all the MLI words and the BSDI test. If we define

$$\begin{aligned} d^e &= min\{MLI_{e_1,2}, MLI_{e_2,2}, MLI_{e_3,2}\}, \\ d^m &= min\{MLI_{m_1,2}, MLI_{m_2,2}, MLI_{m_3,2}\}, \\ d^h &= min\{MLI_{h_1,2}, MLI_{h_2,2}, MLI_{h_3,2}\}. \end{aligned}$$

where e, m, h stand for 'easy', 'medium' and 'difficult'. The variables d^q, $q = e, m, h$ are set to zero if a median child does not know at least one of the three'q' words. In contrast, $d^q = 1$ if the median child knows all of the 'q' words.

Given the latent factor model we estimate, it follows that

$$\begin{aligned}\Pr\left(d^q = 1\right) = &\left[\prod_{j=1}^{3}\left[1-\Phi\left(-\alpha^{MLI}_{q_j,k=1,2}-\beta^{MLI}_{q_j,2}\theta\right)\right]\right]^{\mathbf{1}(d^q=1)} \\ &\left[1-\prod_{j=1}^{3}\left[1-\Phi\left(-\alpha^{MLI}_{q_j,k=1,2}-\beta^{MLI}_{q_j,2}\theta\right)\right]\right]^{\mathbf{1}(d^q=0)}.\end{aligned} \tag{4.4}$$

Let $\hat{\theta}_q$ denote the prediction of the factor θ implied by our latent factor model when $\Pr\left(d^q = 1\right) = 0.5$. Notice that, given the scaling we have described in section 4.3.1, this prediction is measured in terms of the developmental age at which the median child already know the words with difficulty level q:

$$\ln \tau_q = \hat{\theta}_q \tag{4.5}$$

This equation is key for converting the elicited ages under different scenarios to developmental outcomes. We focus on the estimates obtained using equation (4.4), which uses the words in the latent factor model that are included in the beliefs elicitation questions.

4.3.3 Using beliefs data: developmental delays and rates of return to investment

Up to now we have not used the data on elicited beliefs. We now transform maternal answers to measurement of maternal expectation about child development at the end of the period, scaled in terms of the developmental age we defined in section 4.3.2. It is at this point in the analysis that we start using the data from elicited beliefs.

As discussed above, we have four possible scenarios, $s = 1, 2, 3, 4$, and we have three different sets of words in levels of difficulty, $q = e, m, h$. Let $a_{i,s,q}$ denote the age reported by mother i for the set in which the word difficulty-level q and the scenario is s. Define maternal beliefs about the *developmental delay* for scenario s and word difficulty-level q, $d_{i,s,q}$, as follows:

$$d_{i,s,q} = a_{i,s,q} - \tau_q \tag{4.6}$$

For instance, assume that $\tau_e = 21$, so that the median child has already learned the easy words by age 21 months. Suppose, additionally, that mother i states that when $s = 4$ the child will learn all of the "easy" words at age 25 months, so that $a_{i,4,e} = 25$. In this example, mother i's beliefs about *developmental delay* implied by the 4th scenario is 4 months, so that $d_{i,4,e} = 4$. In other words, mother i's beliefs that this hypothetical child is 4 months behind in terms of the median child.

Let τ denote the age at the end of a period considered in the belief's questions We can derive the maternal expectations about end-of-period child development for each scenario as follows:

$$\ln H_{i,1,s,q} = \ln\left(\tau - d_{i,s,q}\right) \tag{4.7}$$

Notice that we have one measure of developmental age for each set of words q (easy, medium and difficult words). These can be seen as different measures of the same theoretical concepts and we use them as such in the last step of our approache.

The expressions in equation (4.7) are used to construct subjective beliefs about the return to investment across different scenarios s. In what follows, we discuss individual perceptions of returns to parental investment and their complementarity with initial conditions. We therefore characterise beliefs about the returns to investment when the initial conditions are high and when they are low. The former can be obtained from equation (4.7) as:

$$r_{i,q}\left(H_0^H\right) = \ln H_{i,1,1,q} - \ln H_{i,1,2,q}, \tag{4.8}$$

while the latter is computed as:

$$r_{i,q}\left(H_0^L\right) = \ln H_{i,1,3,q} - \ln H_{i,1,4,q}. \tag{4.9}$$

Once again we stress that we have multiple measures of these returns, derived from different set of words in terms of their difficulty level.

4.3.4 From answers to beliefs to beliefs about a production function

So far, our approach has mainly scaled data about child development to make it comparable to the questions asked about beliefs and used the latter to compute rates of returns to investment under different initial conditions. In the last step of our approach, we show how we can use $\ln H_{i,1,s,q}$ to estimate maternal beliefs about a production function. Obviously, to perform this last step we need some additional assumptions.

We assume that mothers relate initial conditions and investment to child development (again scaled according to a certain metric) using a specific assumption about the functional form of the production function. In particular, let I_i denote the parent's information set. From the point of view of the parent, we assume that mother i believes that production function of human capital is given by:

$$E\left(\ln H_{i,1}|\, I_i\right) = \mu_{i,0} + \mu_{i,1}\ln H_{i,0} + \mu_{i,2}\ln X_i + \mu_{i,3}[\ln H_{i,0}\ln X_i] + \epsilon_i, \quad (4.10)$$

Notice that such a specification can be estimated on actual data that relate initial conditions and investment to child development. In what follows, we estimate such a function. Whilst the functional form that we fit to actual data is the same as in equation 4.10, in this equation the parameters are specific to each mother in the sample and do not necessarily coincide with the parameters of the production function estimated on actual data. Moreover, as we scaled all the factors involved in the same fashion, the mother-specific parameters μ_i will be comparable to those fitted to actual data. Our analysis, therefore, allows us to compare parental beliefs about the process of human capital development with evidence about the process itself.

We assume that $\ln H_{i,1,s,q}$ is an error-ridden measure of $E\left(\ln H_{i,1}|\, H_{0,s}, X_s\right)$:

$$\ln H_{i,1,s,q} = E\left(\ln H_{i,1}|\, H_{0,s}, X_s\right) + \eta_{i,1,s,q} \quad (4.11)$$

As we mentioned in Section 2, the scenarios we proposed to respondents were chosen after factor analysing the data on investment and initial conditions. Given the IRT models estimated on those data, we can therefore compute a score for investment (X_s) and human capital at the beginning of the period ($H_{0,s}$) corresponding to each scenario. Furthermore, as we said above, we can treat the estimates of expected development obtained with different sets of words (easy, medium and hard) as different measures of the same (unobserved) expected development corresponding to a given scenario. In fact, as we have four different scenarios and three different levels of difficulty, equation (4.11) implies that there are 12 measures of individual expected development. If we substitute equation (4.10) into equation (4.11), we obtain:

$$\ln H_{i,1,s,q} = \mu_{0,i} + \mu_{1,i} \ln H_{0,s} + \mu_{2,i} \ln X_s + \mu_{3,i} \left[\ln H_{0,s} \ln X_s\right] + \eta_{i,1,s,q}. \quad (4.12)$$

The equality requires that we assume that $E\left(\epsilon_i \middle| H_{0,s}, X_s\right) = 0$. This condition means that the scenarios are random with respect to shocks ϵ_i. This assumption is fairly mild because the scenarios are constant across respondents.

Equation (4.12) is a factor model where the μ_i's are the factor and where the factor loadings are known. In particular, the loading for the first factor is 1, for the second factor is $\ln H_{0,s}$, for the third factor is $\ln X_s$, and for the fourth factor is $[\ln H_{0,s} \ln X_s]$. We can therefore estimate the parameters for the distribution of the μ's as well as the variances of the measurement errors. Given these estimates, we can then estimate, for each individual, the vector $\{\widehat{\mu}_{i,0}, \widehat{\mu}_{i,1}, \widehat{\mu}_{i,2}, \widehat{\mu}_{i.3}\}$. These will constitute our estimates of individual beliefs about the production function.

4.4 The Model

We assume parents maximise utility, which depends on household consumption (C_i), the child's human capital at the end of the period, ($H_{i,1}$), and investments (X_i). We allow preferences to depend directly on investment to capture potential psychic costs of investing in children. This formulation, while not standard, also allows for the intervention to affect investments not only through parental beliefs, but also through parental psychic costs. We assume that the utility function is Cobb-Douglas in these three arguments:

$$U\left(C_i, H_{i,1}, X_i\right) = \ln C_i + \lambda_i \ln H_{i,1} + \kappa_i \ln X_i$$

The maximisation problem is subject to a budget constraint. Let P_i and Y_i denote, respectively, the price of investment (relative to the household consumption good) and household income. Because we assume that this is a one-period model, we write:

$$C_i + P_i X_i = Y_i. \tag{4.13}$$

Let $H_{i,0}$ denote the child's human capital at the beginning of the period. Let ϵ_i and ν_i denote, respectively, zero-mean developmental shocks that are known and unknown by the parent at the time that investments are chosen. We assume that the technology of skill formation follows a translog functional form, which is a generalisation of the Cobb-Douglas functional form:

$$\ln H_{i,1} = \delta_0 + \delta_1 \ln H_{i,0} + \delta_2 \ln X_i + \delta_3[\ln H_{i,0} \ln X_i] + \epsilon_i + \nu_i, \tag{4.14}$$

Let Ω_i denote the parent's information set. In this paper, we do not impose the assumption that $\delta = (\delta_0, \delta_1, \delta_2, \delta_3) \in \Omega_i$. Therefore, in our paper, parents do not necessarily observe or know the "true" technology of skill formation but have beliefs about it. From the point of view of the parent, we assume that mother i believes that this technology is:

$$E\left(\ln H_{i,1} | \, \Omega_i\right) = \mu_{i,0} + \mu_{i,1} \ln H_{i,0} + \mu_{i,2} \ln X_i + \mu_{i,3}[\ln H_{i,0} \ln X_i] + \epsilon_i, \tag{4.15}$$

Note the different roles that (4.14) and (4.15) play in the model. Technology (4.14) determines the evolution of the child's human capital, while (4.15) determines parental investment choices.

Finally, let T_i denote the parent treatment status. The maternal information set is $\Omega_i = \left\{H_{i,0}, Y_i, P_i, T_i, \mu_{i,0}, \mu_{i,1}, \mu_{i,2}, \mu_{i,3}\right\}$. The problem of the mother is to maximise expected utility

$$\max_{X_i} E\left[\ln C_i + \lambda_i \ln H_{i,1} + \kappa_i \ln X_i \middle|\, \Omega_i, X_i\right]$$

subject to (4.13), (4.14) and (4.15).

From this problem it is possible to derive an investment equation:

$$X_i = g(P_i, Y_i, H_{i,0}, \epsilon_i; \mu_{i,0}, \mu_{i,1}, \mu_{i,2}, \mu_{i,3}, \gamma, \lambda)$$

The investment equation and the production function (4.14) can be estimated as there are variables in the investment equation that do not enter the production function. However, without further restrictions or information it is not possible to identify separately the μ_i's from the λ and γ. To identify the model, one alternative, is to elicit information on beliefs.

4.5 Identification of the Model

4.5.1 Measurement of Child Development

4.5.1.1 Scales of Child Development

An important step in the identification of the model is to understand how child development (the variables $H_{i,0}$ and $H_{i,1}$ in our model) is measured in the specific study on which we base our analysis. The measurement of child development determines not only the issues that must be confronted to estimate the true technology of skill formation (4.14), but also the parental perceived technology (4.15). It is important to keep in mind that the study is longitudinal, so $H_{i,0}$ is measured at baseline and $H_{i,1}$ is quantified at follow-up, but they are both gauged using the same assessment tools.

The measurement of child development relied on two distinct instruments: the MacArthur Language Inventory (MLI) and the Bayley Scales of Infant Development (BSID). In particular, both $H_{i,0}$ and $H_{i,1}$ are derived using real data (the average of logarithm BSID scales) for the estimation of the true technology of skill formation instead for the estimation of the parental perceived technology we use the scenarios presented to the mother based on the MLI normalised by the BSID scales. In practice, we replace $H_{i,0}$ by $H_{s,0}$ and X_i by X_s where both, $H_{s,0}$ and X_s, were created as hypothetical scenarios in order to elicit beliefs about the perceived technology of skill formation. To get comparable parameters between both production function specifications, we re-scale scenario values according to real data (in terms of mean and variance).

The MLI is divided into two parts that depend on the child's age. MLI I is appropriate for children aged between 12 to 18 months-old. For each word prompted by the interviewer, the parent reports if the child "understands and says the word", "understands, but does not say the word," "neither understands, nor says the word." These possibilities are assigned, respectively, a code number equal to 2, 1, and 0. There are 104 words in the MLI I.

The MLI II is appropriate for children who are at least 19 months-old and at most 30 months-old. In part II, the parents report if the child "says the word" or "does not say the word" that is asked by the interviewer. If the child says the word, then this is assigned a score equal to 1. Otherwise, it is scored equal to 0.

The BSID is a scale that is appropriate for children aged between 1 and 42 months-old. Unlike the MLI, the BSID is based on direct observation of the child by a trained evaluator. The BSID scale produces three scales that will be used in our analysis, a receptive language scale, an expressive language scale and a cognitive scale. All of these are treated as continuous variables. The expressive language scale is the variable that is used to fix the factor loading and normalise the results in terms of the BSID.

4.5.1.2 Item Response Theory

Let the variable θ_i describe child development. We shall make clear below how θ_i relates to the variables $H_{i,0}$ and $H_{i,1}$. Right now, we would like to show how θ_i relates to the scores from the MLI and BSID.

First, note that MLI I is clearly an ordered discrete variable. Thus, we define the index $MLI^*_{i,j,1}$ in the following way:

$$MLI^*_{i,j,1} = \sum_{k=1}^{K} \alpha^{MLI}_{j,k,1} Z_{i,k} + \beta^{MLI}_{j,1} \theta_i + u^{MLI}_{i,j,1},$$

where $u_{i,j,1}^{MLI} \sim N(0,1)$. The variable $Z_{i,k}$ is a matrix of observables variables in dynamic measurement system: constant, male, and age (K=3). The former is adjusted for the age at which the observation i is done but centering it around ages 18 months, so then, a_i, which denote the logarithm of the child's age, is equal to zero if the observation of the MLI I is done at age 18 months. The parameters $\alpha_{j,k=1,1}^{MLI}$ and $\beta_{j,1}^{MLI}$ capture the difficulty and the informativeness of a given word and u_i^j is measurement error. With some assumptions about the distribution in the cross section of the factor θ_i, it will be possible to identify the parameters $\alpha_{j,k,1}^{MLI}$ and $\beta_{j,1}^{MLI}$ as well as the parameters of the distribution of the factor. For instance, we can assume that the factor θ_i is distributed as a mixture of normals.

Let $MLI_{i,j,1} \in \{0,1,2\}$ denote the observed score for child i in word j from the MLI I. The relationship between the score $MLI_{i,j,1}$ and the index $MLI_{i,j,1}^*$ is determined by the following rule:

$$MLI_{i,j,1} = \begin{cases} 0, & \text{if } MLI_{i,j,1}^* \leq 0, \\ 1, & \text{if } 0 < MLI_{i,j,1}^* \leq c_j, \\ 2, & \text{if } c_j < MLI_{i,j,1}^*. \end{cases}$$

where c_j is the cut-off constant in the ordered discrete variable model. Assume, for now, that we observe θ_i. Let Φ denote the CDF of a standard normal random variable. In this case, the contribution to the likelihood of observing score $MLI_{i,j,1}$ for child i in word j is:

$$\begin{aligned} G_{i,j,1}\left(\alpha_{j,k,1}^{MLI}, \beta_{j,1}^{MLI}, \theta_i\right) &= \left[\Phi\left(-\alpha_{j,k,1}^{MLI} - \beta_{j,1}^{MLI}\theta_i\right)\right]^{\mathbf{1}(MLI_{i,j,1}=0)} \times \\ &\left[\Phi\left(c_j - \alpha_{j,k,1}^{MLI} - \beta_{j,1}^{MLI}\theta_i\right) - \Phi\left(-\alpha_{j,k,1}^{MLI} - \beta_{j,1}^{MLI}\theta_i\right)\right]^{\mathbf{1}(MLI_{i,j,1}=1)} \times \\ &\left[1 - \Phi\left(c_j - \alpha_{j,k,1}^{MLI} - \beta_{j,1}^{MLI}\theta_i\right)\right]^{\mathbf{1}(MLI_{i,j,1}=2)}. \end{aligned}$$

Second, note that MLI II is a binary discrete variable. Thus, we define the index $MLI_{i,j,2}^*$ in the following way:

$$MLI_{i,j,2}^* = \sum_{k=1}^{K} \alpha_{j,k,2}^{MLI} Z_{i,k} + \beta_{j,2}^{MLI}\theta_i + u_{j,2}^{MLI},$$

where $u_{i,j,2}^{MLI} \sim N(0,1)$ .Let $MLI_{i,j,2} \in \{0,1\}$ denote the observed score for child i in word j from the MLI II. It follows that:

$$MLI_{i,j,2} = \begin{cases} 0, & \text{if } MLI^{*}_{i,j,2} \leq 0, \\ 1, & \text{if } MLI^{*}_{i,j,2} > 0. \end{cases}$$

Therefore, the contribution to the likelihood can be written as:

$$G_{i,j,2}\left(\alpha^{MLI}_{j,k,2}, \beta^{MLI}_{j,2}, \theta_i\right) = \left[\Phi\left(-\alpha^{MLI}_{j,k,2} - \beta^{MLI}_{j,2}\theta_i\right)\right]^{\mathbf{1}(MLI_{i,j,2}=0)} \times \left[1 - \Phi\left(-\alpha^{MLI}_{j,k,2} - \beta^{MLI}_{j,1}\theta_i\right)\right]^{\mathbf{1}(MLI_{i,j}}$$

Third, the BSID is a continuous variable. Thus, let $BSID_{i,j}$ denote the observed score for child i in language module j. The relationship with the variable θ_i is captured by:

$$BSID_{i,j} = \sum_{k=1}^{K} \alpha^{BSID}_{j,k} Z_{i,k} + \beta^{BSID}_{j}\theta_i + u^{BSID}_{i,j},$$

where $u^{BSID}_{i,j} \sim N\left(0, \sigma^2_j\right)$. Therefore, the contribution to the likelihood is:

$$G_{i,j,3}\left(\alpha^{BSID}_{j,k}, \beta^{BSID}_{j}, \theta_i\right) = \left(\frac{1}{\sigma_j}\right)\left(\frac{1}{\sqrt{2\pi}}\right) \exp\left\{\frac{\left(BSID_{i,j} - \alpha^{BSID}_{j,k} - \beta^{BSID}_{j}\theta_i\right)^2}{2\sigma^2_j}\right\}.$$

Fourth, define the variable $\chi_{i,j} = 1$ if the j^{th} observation for child i is from MLI I, $\chi_{i,j} = 2$ if the j^{th} observation for child i is from MLI II, and $\chi_{i,j} = 3$ if the j^{th} observation for child i comes from BSID. Let $\alpha_j = \left(\alpha^{MLI}_{j,1}, \alpha^{MLI}_{j,2}, \alpha^{BSID}_{j}\right)$ and define β_j in the same fashion. The contribution to the likelihood from item j and child i is:

$$\begin{aligned} G_{i,j}\left(\alpha_j, \beta_j, \theta_i\right) \quad = \quad & \mathbf{1}\left(\chi_{i,j} = 1\right) \times G_{i,j,1}\left(\alpha^{MLI}_{j,k,1}, \beta^{MLI}_{j,1}, \theta_i\right) + \\ & \mathbf{1}\left(\chi_{i,j} = 2\right) \times G_{i,j,2}\left(\alpha^{MLI}_{j,k,2}, \beta^{MLI}_{j,2}, \theta_i\right) + \\ & \mathbf{1}\left(\chi_{i,j} = 3\right) \times G_{i,j,3}\left(\alpha^{BSID}_{j,k}, \beta^{BSID}_{j}, \theta_i\right). \end{aligned}$$

Fifth, the contribution of child i to the likelihood is:

$$L_i\left(\alpha_k, \beta, \theta_i\right) =_{j=1}^{J} G_{i,j}\left(\alpha_{j,k}, \beta_j, \theta_i\right), \tag{4.16}$$

where $\alpha_k = \left(\alpha_{j,k}\right)_{j=1}^{J}$ and β is defined in a similar way.

Finally, the likelihood takes into account the fact that θ_i is not observed for any child. Therefore, we must integrate out the distribution of θ:

$$L_i(\alpha_k, \beta) = \int_{j=1}^{J} G_{i,j}(\alpha_{j,k}, \beta_j, \theta_i) f(\theta) d\theta, \tag{4.17}$$

This implies that the likelihood function is:

$$L =_{i=1}^{N} \int L_i \tag{4.18}$$

We can estimate the parameters of the IRT model (the parameters α and β and the distribution of θ) by maximising the likelihood (4.18). Given estimates for α_k and β, we can estimate a child-specific θ_i by maximising the likelihood (4.17) for each child. Figure 4.1 displays the estimated density of the factor scores of child development (θ).

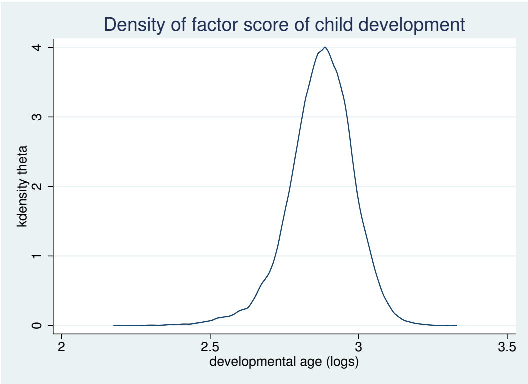


Figure 4.1: Factor Score

4.5.2 Estimation of the Technology of Skill Formation and Utility Function

Now that we have estimated the parameters that describe mean beliefs, we show how they can be helpful in the estimation of the parameters of the technology of skill formation and the utility function. We start by proposing a novel way to identify the true parameters of the technology of skill formation. Let $\Delta_{k,i} = (\delta_k - \mu_{k,i})$ for $k = 0, ..., 3$. Consider the difference between equations (4.14) and (4.15):

$$\ln H_{i,1} - E\left(\ln H_{i,1} \middle| \Omega_i\right) = \Delta_{0,i} + \Delta_{1,i} \ln H_{i,0} + \Delta_{2,i} \ln X_i + \Delta_{3,i} \left[\ln H_{0,s} \ln X_s\right] + \nu_i.$$

In particular, note that the error term ν_i is not in the parent's information set. This observation suggests that a method of moments procedure that explores the following moment conditions:

$$E\left\{\ln H_{i,1} - E\left(\ln H_{i,1} \middle| \Omega_i\right) - \Delta_{0,i} - \Delta_{1,i} \ln H_{i,0} - \Delta_{2,i} \ln I_i - \Delta_{3,i} \left[\ln H_{0,s} \ln X_s\right] \middle| \Omega_i\right\} = 0.$$

Next, we consider the estimation of the parameters of the utility function. Here, we explore the information from the first-order condition:

$$E\left\{ I_i - \frac{\kappa_i + \lambda_i \mu_{i,2} + \lambda_i \mu_{i,3} \left[\ln H_{0,s} \ln X_s\right]}{1 + \kappa_i + \lambda_i \mu_{i,2} + \lambda_i \mu_{i,3} \left[\ln H_{0,s} \ln X_s\right]} \frac{Y_i}{P_i} \middle| \Omega_i \right\} = 0$$

Remember that T_i is the parent's treatment status. In our empirical analysis, we further assume that:

$$\begin{aligned} \lambda_i &= \lambda_0 + \lambda_1 T_i \\ \kappa_i &= \kappa_0 + \kappa_1 T_i \end{aligned}$$

4.6 Data and Empirical Results

4.6.1 Data

We designed a module to elicit parental beliefs about the process of child development within the second follow up to a sample of mothers of 1200 children who were part of a early childhood intervention through a Randomized Controlled Trial in Colombia. The aim is to elicit beliefs about the production function, *not the ability of their children*. The basic structure was guided by the Jamaica experiment by Sally Grantham-McGregor [Grantham-McGregor et al. (1991)], designed to evaluate the effect of two different interventions, the first being a stimulation intervention delivered through weekly home visits to stimulate the child and involve the care giver and child in a number of structured visits, while the second was micronutrient supplementation. The intervention use the infrastructure of an existing welfare programme and community women to deliver it. The Jamaica curriculum was adapted to the Colombian context and aim to promote child-development in an integrated manner (motor, language, cognitive, socio-emotional skills), it also encourage mothers to teach her children based on events surrounding daily routine activities. In particular, the curriculum was based on picture books, pictures to stimulate conversation, puzzles, cubes/blocks and patters, toys from recycled materials and language games and songs. It targeted 1,429 children aged 12-24 months at baseline in 96 semi-urban towns. The design was a 2x2, so children were randomised into 4 groups (at the village-level): Stimulation, Supplementation (micronutrients), Stimulation + Supplementation and Control.

That intervention had some impacts on a variety of outcomes. The impacts of these interventions are described in [Attanasio et al. (2014)]. This evaluation work was based on two surveys on the children in the study and their mothers: the baseline survey collected before the intervention started in 2009-2010 (children aged 12-24 months) and one after 18 months at the end of 2011. Several measurement on children development were collected (Motor and Cognitive Development: Bailey Test, Socio-emotional Development: Bates Temperament, Language Development: MacArthur-Bates, Height, weight, haemoglobin and Morbidity, Food Intakes, Child care arrangements and Time Use, among others) as well as information on mothers and families using a general household survey (SES, Education, Labour supply, Time use, Reproductive history, Health conditions, Depression, Knowledge on Parenting, Parenting Practices and the Home Environment, among others).

About two years after the end of the intervention, the mother and the children that participated into the study were contacted again. The third data collection, that happened in the fall of 2013, included, among other things, a module to elicit parental beliefs.

Using the data collected we can get some idea about returns to investment, Figure 4.2 shows results to the medium words. The red density refers to the rate of returns conditional on high initial conditions, while the blue line to that conditional on low initial conditions, mothers perceive the returns to maternal investment to be slightly higher under low initial conditions than in the alternative scenario, which means, under low initial conditions and high maternal investment a child would say medium words earlier in age than a child with low initial conditions and low maternal investment.

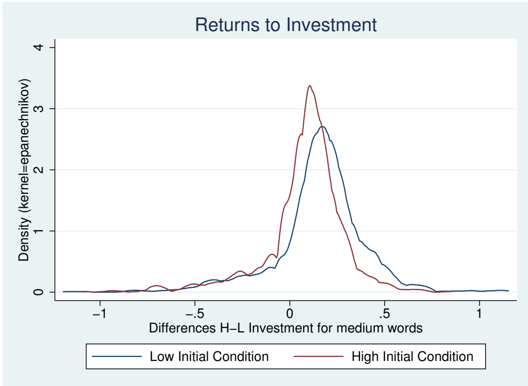


Figure 4.2: Returns to Investment

4.6.2 Empirical Results

In this section, we describe the data on parental beliefs we have elicited and show how they translate into beliefs about the returns to parental investment following the procedure described in Section 4.2. In particular, we describe the distribution of individual beliefs, how they relate to socioeconomic status and how returns to investment are perceived to vary with initial conditions. Finally we relate beliefs to individual behaviour and show whether they have been affected by a parental intervention.

4.6.2.1 Beliefs about the Returns to Investments

Each mother reports her beliefs about the the language development corresponding to each of the various scenarios. Using equation (4.7) and the procedure discussed above, we translate the answers given into a developmental age corresponding to each of the scenarios considered. As we have three different estimates of the perceived developmental age for each investment/initial condition scenario (corresponding to easy, medium and difficult words), we consider their average.

In Figure 4.3, we plot the four distributions of the (log average) developmental ages corresponding to the four scenarios. If respondents understood the question, the distribution of developmental ages should move to the right as we move from the worst scenario (low initial conditions and low investment) to the best one (high initial conditions and high investment). Analogously, when moving from the lowest scenario one of the two intermediates, we should also observe a movement to the right of the density.

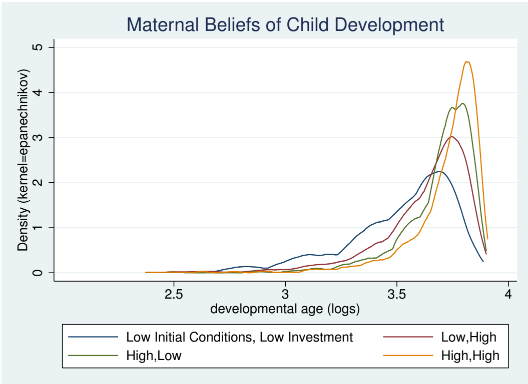


Figure 4.3: Maternal Beliefs of Child Development

The results we obtain are consistent with these predictions. When moving from the worse to the best scenarios the distribution moves to the right and, interestingly, becomes more less disperse. Of course, the predictions for the intermediate cases are ambiguous as they will depend on the degree of substitutability or complementarity between investment and initial conditions as perceived by respondents.

In Figure 4.4, we plot the returns to investment computing them in equations (4.8) and (4.9), as the differences of (log averages) developmental ages corresponding to high and low investment. The red density refers to the rate of returns conditional on high initial conditions, while the blue line to that conditional on low initial conditions, The striking feature of this Figure is that mothers perceive the returns to parental investment to be much higher (although more disperse) under low initial conditions than in the alternative scenario.

Table 4.2 relates the subjective expected returns to investment for low and high initial conditions to socioeconomic characteristics of the mother who expressed those beliefs. In particular, we regress perceived returns (under low and initial conditions) on age, on two education dummies (indicators for

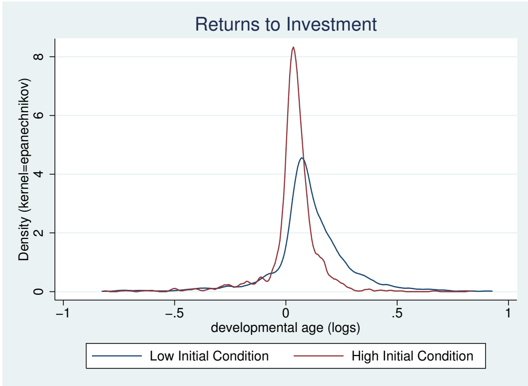


Figure 4.4: Subjective Expected Returns

primary or secondary education, with the no-education being the excluded group), the CES-D index of depression, the score in the Raven progressive matrices test and an indicator of whether the child targeted by the intervention is male. Of these variables, the only one that is significant in both regressions is the score in the Raven tests, indicating that women with higher Raven tests have higher expected returns to maternal investment, both for low and high initial conditions.

Table 4.2: Returns on Investment and SE characteristics

	Low Initial Cond.	Low Initial Cond.	High Initial Cond.	High Initial Cond.
Mother's age	0.002	0.002	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Education (primary)	0.131	0.131	-0.002	-0.002
	(0.089)	(0.089)	(0.055)	(0.055)
Education (secondary and more)	0.125	0.125	0.016	0.016
	(0.089)	(0.089)	(0.055)	(0.055)
CES-D	0.001	0.001	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Raven test	0.002^{**}	0.002^{**}	0.002^{***}	0.002^{***}
	(0.001)	(0.001)	(0.000)	(0.000)
=1 if TC male	-0.020^{*}	-0.019^{*}	-0.006	-0.006
	(0.010)	(0.010)	(0.008)	(0.008)
Treatment		0.007		0.003
		(0.012)		(0.009)
R^2	0.018	0.018	0.032	0.032
F	2.581	2.206	4.227	3.627
Observations	1161	1161	1161	1161

Standard errors (in parentheses) are clustered at municipality level, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Are Beliefs about Returns Affected by Treatment? [Attanasio et al. (2014)] have shown that the intervention evaluated by the data we are using had an impact on several measures of child development, including cognition and language. [Attanasio et al. (2015)] analysed the data from the same intervention and showed using a production function for cognitive and non-cognitive skills with factors of material investments and time investments by parents that the intervention shifts significantly the distribution of the two investment factors. Using our survey for collecting beliefs, we can plot subjective expected returns to investment for low and high initial against the treatment variable. Figure 4.5 shows the impact of the intervention on the Subjective Expected Returns, confirming the results shown in Table 4.2.

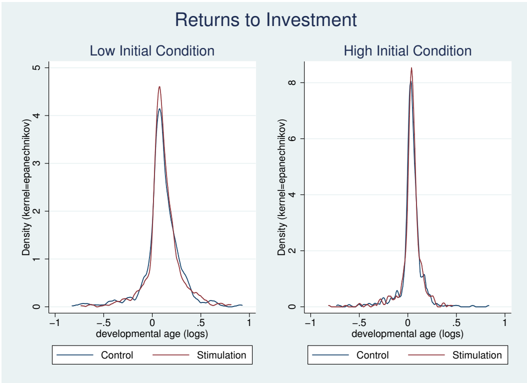


Figure 4.5: Subjective Expected Returns and Intervention

Do Beliefs about Returns Correlate with Investments? Having estimated parental beliefs on returns to investment we want to relate them to investment. We start by considering a reduced form equation where investment is regressed on its determinants, *including perceived returns*. We consider separately returns conditional on "low" initial conditions and "high" initial conditions. We have three measures of these returns, one corresponding to each set of words (easy, medium and hard). We can use one measure of expected return and instrument it with the other two, to allow for measurement error. Table 4.3 shows that maternal beliefs about high initial conditions correlate with maternal investments. A different and more efficient way to answer this could be through the estimation of a factor model.

Table 4.3: Investment and Returns on Investment

	OLS_average	OLS_hard	IV
Return to Low Initial Condition	-0.306*	-0.269**	-0.235
	(0.178)	(0.126)	(0.178)
Return to High Initial Condition	0.884***	0.386**	0.930***
	(0.250)	(0.191)	(0.249)
R^2	0.009	0.005	.
F	6.820	3.909	7.012
Observations	1200	1200	1200

Standard errors (in parentheses) are clustered at municipality level, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.6.2.2 Beliefs about the Technology of Skill Formation

The mother's perceived production function is defined by equation (4.15). Our procedure yields estimates of the coefficients for each mother. Figure 4.6 shows the density of the beliefs for each factor in the mother's perceived production function. The estimates for the initial condition is less volatile and more concentrate than the estimate for maternal investment. These inputs are created using the scenarios presented in the beliefs survey.

Table 4.4: Maternal Investment

Maternal Investment	$\ln X$
Low	-3.2103
High	1.0610

Table 4.5 shows the median of the perceived production function as well as the estimates of the 'true' production function. The difference between both production functions are mainly due to the investment factor definition in each case. For the 'true' production function the data used is consistent with the data collected about maternal investment during the first follow up, instead, the 'perceived' production function uses the data collected through the module to elicit parental beliefs, in particular, Table 4.4 shows the levels of maternal investment for each case derived from the IRT estimation.

Table 4.5: Production Function Estimates: Perceived Median and "True"

	Perceived		"True"
μ_0	2.433	δ_0	2.362
			(0.107)
μ_1	0.454	δ_1	0.418
			(0.037)
μ_2	0.197	δ_2	0.414
			(0.132)
μ_3	-0.065	δ_3	-0.132
			(0.047)

Standard errors (in parentheses) are clustered at municipality level

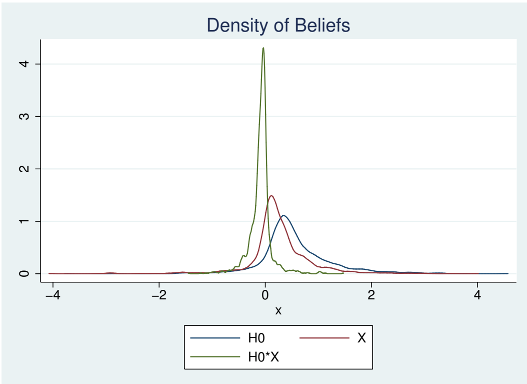


Figure 4.6: Perceived production functions

Figure 4.7 shows the marginal product of investment used in each production function. The blue line is the marginal productivity for the objective production function which is higher than the marginal productivity that is perceived by the mothers in their production function, and that is actually related to the estimates for μ_2 in Table 4.5.

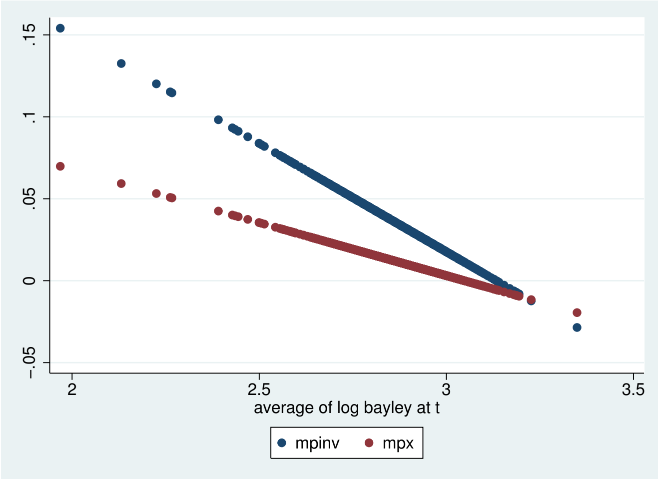


Figure 4.7: Marginal Product of Investments

Are Beliefs about Technology Affected by Treatment? Table 4.6 shows how beliefs about the TFP and initial conditions are affected by the treatment, positively and negatively respectively. Mother's age, maternal education, maternal socioemotional skills based on the measure of depression using the CES-D test, maternal cognitive skills based on the Raven test and a dummy variable equal to 1 if the child is male do not show any correlation with the beliefs recovered from the subjective production function.

Table 4.6: Production Function Estimates and SE characteristics

	μ_0	μ_0	μ_1	μ_1	μ_2	μ_2	μ_3	μ_3
Mother's age	-0.013	-0.014	0.005	0.005	0.002	0.002	-0.001	-0.001
	(0.010)	(0.010)	(0.004)	(0.004)	(0.004)	(0.004)	(0.001)	(0.001)
Education (primary)	1.550	1.550	-0.534	-0.534	0.453	0.453	-0.157	-0.157
	(0.990)	(1.002)	(0.346)	(0.350)	(0.308)	(0.309)	(0.108)	(0.108)
Education (secondary and more)	1.455	1.460	-0.494	-0.495	0.375	0.375	-0.128	-0.129
	(0.983)	(0.994)	(0.343)	(0.347)	(0.313)	(0.314)	(0.110)	(0.110)
CES-D	0.004	0.006	-0.001	-0.002	0.005	0.005	-0.002	-0.002
	(0.011)	(0.011)	(0.004)	(0.004)	(0.004)	(0.004)	(0.001)	(0.001)
Raven test	0.000	0.000	0.001	0.001	-0.001	-0.001	0.000	0.000
	(0.008)	(0.008)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)
=1 if TC male	0.132	0.141	-0.043	-0.046	-0.049	-0.048	0.016	0.016
	(0.109)	(0.109)	(0.038)	(0.038)	(0.036)	(0.036)	(0.013)	(0.013)
Treatment		0.249**		-0.084**		0.014		-0.005
		(0.116)		(0.040)		(0.035)		(0.012)
R^2	0.010	0.014	0.010	0.013	0.014	0.014	0.014	0.014
F	0.927	1.316	0.973	1.344	2.039	1.760	2.195	1.890
Observations	1161	1161	1161	1161	1161	1161	1161	1161

Standard errors (in parentheses) are clustered at municipality level, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.7 Conclusions and further work

There is a demand in shifting the research focus, from understanding the impacts of resources, to understanding the determinants of behaviours. We attempt to address these questions by analysing maternal beliefs on an impact evaluation of a parenting program in Colombia using Randomized Control Trials methods. The intervention is likely to change three dimensions of parental beliefs: beliefs about what is the best way to raise children, beliefs about how sure parents feel about the parenting task, and parental perception about the returns to investments in children. By collecting pre and post-treatment data from a sample of more than 1,300 households, we investigate the effects of these beliefs on parental investments and child development.

Our project has several innovations. First, this is the first work collecting several measures of beliefs that complement each other, forming a much more complete set of beliefs than ever considered before. Second, we use the exogenous variation of beliefs provided by the intervention to investigate treatment effects controlling for measurement error. Third, we investigate the indirect effects of beliefs on outcomes by changing parental investments and the direct effects of changing the productivity of those investments under relatively weak assumptions. Fourth, we take advantage of the experimental data to identify separately heterogeneous preferences and expectations in the estimation of a model of parental investments in children in which we emphasise that parents are uncertain about the returns to investments in the technology of skill formation. Such a model will be potentially useful for the simulation of more cost-effective early child development policies.

Evidence from the data is encouraging about the potential findings. First, we have elicited maternal beliefs about the production function. Second, we have shown how to relate answers about developmental age under different scenarios to beliefs about returns to investment and parameters of the production function. Third, we find that parents think that the productivity of investment is much higher for low initial conditions than higher initial conditions. Finally, we want to extend this approach and estimate simultaneously the production function, the perceived production function and the investment strategy.

Chapter 5

A Dynamic Model of Early Parental Investments in Children's Human Capital

5.1 Introduction

Recent research demonstrates that the effects of early childhood environments last a lifetime. There is an extensive literature on how parental characteristics and household environment affect investment in children's human capital, but little about how parents' investment decisions and the structure of family dynamics behave. The pathways linking parental characteristics to long-term child outcomes remain unclear. A better understanding of these relationships requires novel modes of inquiry that transcend those of any particular discipline.

Numerous studies establish that measured cognitive ability is a strong predictor of economic success in life (summarised in Cawley et al., 2001) and that non-cognitive ability are likely to be an essential determinant of social success in life (Bowles and Ginties, 1976) and for predicting wages, schooling, and participation in risky behaviours (Heckman and Rubinstein, 2001). As a result, there is a vast literature on how parental characteristics and household environment affect investment in children's human capital, but there is little information about how parents' investment decisions, behave: What are the channels, if any, through which these decisions affect child outcomes? Do these decisions, respond to incentives/stimulation? Do public policies influence child outcomes via these decisions?

Structural economic models of individual behaviour, when correctly identified, can lead to estimated policy-invariant parameters that govern preferences or technologies and can be used to learn about behaviours and therefore to evaluate the effectiveness of a program even when that policy has not been implemented. Moreover, structural models provide the tools to understand the mechanisms behind observed decisions in a way that is consistent with accepted theories of economic behaviour. They also offer the flexibility to accommodate observed and unobserved heterogeneity explaining behaviours. Recently, the most advanced research in the field has turned to the credible estimation of structural models that can inform social policies in more comprehensive ways about relevance, heterogeneous impacts, and the optimal timing of programs.

Even though we know that there is a positive return on investment in early years almost all societies still face gaps in childhood development. Unfortunately, inequality perpetuates poverty in most developing countries due to the lack of efficient public policies, so there is a high return of implementing optimal policies for flourishing human capital skills in early years by improving the conditions for disadvantaged children. The mechanisms generating family influence and long-term child outcomes remain unclear. My goal is to understand the relevance of these parental decisions and family dynamics in child outcomes. Research questions that I seek to answer are: If the return on investment is higher for providing more resources to families or instead of improving their parenting skills? What parents are doing/how they behave? These questions are still unanswered and therefore is crucial to comprehend which are the mechanisms parenting enhance the formation of early human capital for its multiple dimensions.

I aim to expand the scope of research on child development to explicitly account for the dynamic interpersonal relationships of attachment, interaction, and scaffolding emphasised in the literature on early child development as well as the fact that it is indispensable to develop more complex economic analysis where preferences, technologies, parental decisions and the importance of dynamics are simultaneously considered and estimated in a model. Using models as the ones that I develop in this chapter it is possible to understand the mechanisms behind decision-making and to simulate policies ex-ante that are crucial to addressing all these questions.

The objective of this chapter is to go beyond the analysis of how constraints impact parental investments in children, and to understand how parenting, in particular, the attitudes towards child-rearing, determines investments in children and their subsequent development. To do so, I integrate a children's human capital model with multiple stages of childhood into a dynamic framework to explain parental investment decisions and children's technology skill formation accounting for unobserved heterogeneity. Parental Investment in children can only be adequately understood when preferences, technologies and choice sets are simultaneously considered. One of the most important aims of this chapter is to explore synergies from both early childhood development and family interactions by analysing how resources are distributed within household and investments in human capital. Therefore, the aim is to identify through what mechanisms parental investment is modified and understand how policies that change the incentives of parental investments affect the early childhood development of their children taking into account parental preferences, unobserved heterogeneity, initial endowments, constraints and parental choices.

For doing so, in this work in progress, I exploit the rich panel structure of the *Encuesta Longitudinal de Primera Infancia* (Early Childhood Longitudinal Survey (ELPI)) survey in Chile. It is a representative sample of the country and provides detailed information about socioeconomic characteristics, home assessment, and cognitive and non-cognitive tests for around 15,000 children, and the objective is twofold: First, identify through what mechanism parental investments are modified answering questions, such as What do parents (not) do? What constraints do they face? What are the key pathways between the early childhood dimensions and cognitive and non-cognitive skills? How are resources allocated? Second, understand how policies that change the incentives of parental investments affect the abilities of their children, and, which social programs change the incentives of individuals to acquire skills and fulfil capacities and modify their outcome and those in their influence, including their children.

I contribute to the literature of child development in five important ways: first, I am modelling a child production function jointly with a real life cycle model taking account of parental constraints, father choices and unobserved heterogeneity (income shocks). Second, I include a quality measure of parental time with the child. Third, I add choices of parental investments not only in a wide range of cognitive skills but also in non-cognitive skills, fourth, I use unique data for developing countries while previous studies used developed countries datasets (as the Panel Study of Income Dynamics (PSID) and the National Longitudinal Survey of Youth (NLSY)), and fifth, I include assets in the model which implies it is possible to study precautionary saving, credit constraints and imperfect markets. This model also offers the possibility to compare two levels of investments, the one they are actually doing through the estimation of the productivity of investment in the production function and the optimal decision through the budget constraint. The implication of difference between both investments is that then one of them is misrepresented which implies distorted beliefs.

The solution and estimation of the model implies hard work, but only capturing both parent choices, plus unobserved heterogeneity can one reasonably expect to answer policy questions that bear upon how households will change their behaviours in the amount of monetary and time investment in response to changes in policies such as parental leave, child care facilities and, parenting programs. In this chapter, I present the model that it is currently being solved and it will be estimated for evaluating counterfactuals.

5.2 Related Literature

Early Childhood Development has been in the centre of the debate in the literature over the last years given its implications for welfare society. An increasing body of studies in neuroscience, psychology and economics, shows that first years of life are critical for future development of children (Thompson, (2001); Van der Gaag, (2005); Noble et al. (2007); Crawford et al. (2010)). In particular, Heckman et al. (2006) find that a low dimensional vector of cognitive and non-cognitive abilities explain a variety of labour market and behavioural outcomes. Cognitive abilities are the child's ability to learn and solve problems, for example, explore the environment with their hands or eyes or to resolve math problems, and moreover to incorporate the understanding of the language, on the other hand, non-cognitive abilities are the socioemotional abilities. Stimulation at early years of these abilities plays therefore a crucial role in the child's future and the social development of future generations. As a result, there is a large literature on how parental characteristics and household environment affect investment in children's human capital, but there is little information about how parents' investment decisions behave: What are the channels, if any, through which parents' decisions affect child outcomes? Do their decisions respond to incentives/stimulation? Can parents' decisions/behaviour be affected through public policies and by doing so change child outcomes?

Some of these questions have been addressed using different models and methodologies, but stylised facts have not yet emerged. The literature has reported a strong correlation between parental investment and children's human capital using a reduced-form approach of a production function for cognitive achievement as summarised in Todd and Wolpin (2003, 2007). However, as proposed by Cuhna et al. (2007, 2010), it is essential to building a model of skill formation with multiple stages of childhood to account for a large body of evidence in Early Childhood Development.

Until now, most of the research has been based on reduced form, and static approach using production functions for cognitive skills assuming inputs are perfect substitutes (Todd and Wolpin, 2003, 2006), roughly a third report positive effects of maternal employment on child cognitive outcomes, a third report negative, and the remaining report either insignificant effects or effects that vary depending on the group studied. Therefore, the economic literature has stressed the importance of disentangling the mechanisms behind parental investment, Heckman et al. (2007 and 2010) develop a dynamic structural model with multiple periods for modelling production functions where inputs are complements at different stages using linear and non-linear production functions finding optimality in life-cycle with monetary investment but without incorporating labour supply. The second wave of papers are also based on dynamic structural modelling but incorporating labour supply into the analysis. Liu et al. (2010) study migration and mother's labour supply, Bernal (2008) shows the effect of child care decisions in the maternal labour supply, Gayle et al. (2006) study mother's labour supply and fertility decisions, meanwhile, Gayle (2011) and Del Bocca et al. (2014) are the first ones to incorporate both parents into the analysis. Even though several studies have emphasised the strong interactions between time spent with the child and monetary investment in the child, to date relatively few studies have explicitly modelled both decisions jointly. Moreover, few have made due allowance for unobserved heterogeneity, and none of these studies model the parameters of a dynamic framework capable of capturing the inherently dynamic nature of parental decisions and parental investment.

A common limitation of previous studies is that they have failed to control for potential endogeneity problem fully: parents that work may be systematically different from parents who do not work, and the child's skills itself may influence parental decisions of whether to work or not, moreover, parents are heterogeneous in their skill endowments, the constraints they face, and their tastes, therefore is crucial to allow parental decisions to depend on these unobserved heterogeneous characteristics of both parents. I develop a model of parental choices jointly with a child production function for cognitive and socioemotional abilities. This type of model allows implementing a correction for endogeneity in the sense that we can adjust for the fact that certain types of parents are more likely to invest in their children. Most importantly, the model can be used to assess the effects of counterfactual policy experiments (policies related to parental leave, child care subsidies, and other incentives for parents to stay at home).

This paper differs from previous research in five important ways: I develop a child production function jointly with a realistic life cycle model that includes parental constraints, fathers choices, unobserved heterogeneity, quality measure of time with the child for both parents, add choices of parental investments not only in a wide range of cognitive skills but also in socioemotional skills, plus including a measurement system to recover the underlying factors in the estimation of the production functions, and I use a new data for developing countries instead of using PSID and NLSY surveys.

5.3 Data and Preliminary Evidence

One of the major limitations for disentangling the determinants of childhood development in Latin America and particularly in Chile is the lack of high quality and detailed data on children under 5 years old. In this section, I present new data that allows researchers to assess the impact of early childhood policies and to provide valuable information for the evaluation and design of social policies in this field since this data contains variables determining children's cognitive and non-cognitive abilities.

5.3.1 Survey Description

The Early Childhood Longitudinal Survey (ELPI) is a rich longitudinal panel dataset from Chile. Chile is the country of Latin America with the highest GDP per capita however, within Chile a high degree of inequality persists in most dimensions of human resources, as well as in income and wealth, with considerable human resource gaps across income groups that start very early in the life-cycle (e.g., around 0.8 standard deviation in language scores for 3-5 year olds between top and bottom quartiles) and tend to persist or even increase over the life cycle.

The first wave was collected in 2010 that covers a representative sample of Chilean households with one child between 0 and 5 years old in urban and rural areas. The target population are children who were born on January 1, 2006, and August 31, 2009. The sample includes different cohorts of children, distinguished by year of birth.[1] The second wave was collected in 2012 and included around 85% of the initial sample wave, but it also incorporated information about a refreshing sample adding cohorts for children between 0 and 2 and a half years old, this target population are children who were born on September 1, 2009, and December 31, 2011 (total 28 months).

The survey contains two major areas. The first area is a questionnaire divided into two main sections. The first section includes questions designed for each of the household members, while the second one contains questions

[1] The sampling frame corresponds to 1.297.822 birth records from March 1, 2004, and August 31, 2009, however the final sampling frame used information from January 1, 2006, and August 31, 2009 (total 44 months) due to the test's timing.

that apply only to the person that answer the questions, i.e., the mother of the target child. The standard Socioeconomic household questionnaire includes information about educational background, labour force participation, income, childcare use, information on female empowerment and household decision-making, interactions between parents and the child and detailed information of investment in the child among others.

The second area of the survey collects a set of cognitive, socioemotional and executive function tests for both the target child and the mother, as well as anthropometrics measures. There is also information about home assessment using the Home Observation for Measurement of the Environment (HOME) Inventory and the Family Care Indicators (FCI). In the next subsection, there is more information about these measures.

5.3.2 Measurements

Tests applied in both ELPI measured the development of children in different areas, such as motor, cognitive, language, emotional and social areas. Evaluate the overall development of children allows us to identify the areas in which they have higher and lower achievements which are relevant at the time of reporting social and educational policies. For all of these measurements, we will identify two underlying factors to explain cognitive and socioemotional skills for both, the mother and the child. The factors for the child will be observed two times t and $t+1$ while the ones for mothers will be observed only at t and we will assume these are a stable measure of their intrinsic skill levels.

Measures of child development outcomes - cognitive:

Scale of Psychomotor Development Evaluation (EEDP) EEDP is a Chilean instrument that, as mentioned by the authors, is used to measure performance against certain situations to be resolved to require a certain degree of psychomotor development. It is administered from 6 months to 23 months with 30 days. It consists of 75 items, five items for each month of age. Four relatively independent and specific operating areas have distinguished themselves within the process of psychomotor development: Motor that includes gross motor, general and specific coordination, including postural reactions and locomotion. Language that covers both verbal and non-verbal language, such as reactions to sound, soliloquy, vocalisations, comprehension and verbal emissions. Social that refers to the child's ability to react to people and to learn through imitation and Coordination that includes the child's reactions that require coordination of different functions. Used in ELPI 2010.

Battelle Developmental Inventory (BDI 2010) It defined as a battery to evaluate basic cognitive abilities in children, applied on an individual basis from 6 to 23 months 30 days. It is focused on five areas (De la Cruz and Gonzalez, 1998): Personal / Social that evaluates the capabilities and features that allow the child to establish meaningful social interactions. Valued behaviours are grouped into six sub-areas (interaction with adults, expression of feelings/affection, self-concept, interaction with peers, collaboration and social role); Adaptive that appreciates the ability of children to use the information and skills assessed in the other areas. It evaluates self-help skills and tasks that require these skills. The first are the behaviours that allow the child to become more independent in feeding, dressing and grooming. The latter represent the capacity of children to pay attention to be specific stimuli during periods getting longer, to take personal responsibility for their actions and initiate activities for a specific purpose, acting appropriately to complete them. The conducts of this area are grouped into five sub-areas (attention, food, dress, personal responsibility and grooming); Motor that evaluates gross motor development and the capacity of children to use and control the muscles of the body (gross and fine motor development). Behaviours appreciated in this area are grouped into 5 sub-areas (muscular control, body coordination, locomotion, fine motor and perceptual motor skills); Communication where there are a reception and expression of information, thoughts and ideas through verbal and nonverbal means. This area is divided into two main sub-areas (receptive communication and expressive communication); and Cognitive which are appreciate skills and conceptual abilities. These behaviours are grouped into four sub-areas (perceptual discrimination, memory, reasoning and skills school and conceptual development). Used in ELPI 2010.

Psychomotor Development Test (TEPSI) TEPSI is a Chilean screening instrument, i.e., a thick evaluation that allows to know the level of performance in terms of psychomotor development of children between two and five years relative to a statistical norm established by age group, and determine whether this performance is normal, or is under expected through the observation of the child?s behaviour in situations proposed by the examiner. It is administered to children from 24 months to 35 months with 30 days. It measures three basic areas of child development (coordination, language and motor). Used in ELPI 2010.

TVIP Test de Vocabulario en Imágenes de Peabody/ PPVT Peabody Picture Vocabulary Test The TVIP is the result of the Spanish adaptation of the Test Peabody Picture Vocabulary, of massive use in English-speaking countries. Both versions are used for educational, research and clinical purposes. It is a psychometric test that measures an individual's receptive or hearing vocabulary. It is administered to children between 30 and 60 months. Used in both ELPI 2010 and 2012.

Battelle Developmental Inventory Screening Test, 2nd ed. (BDI 2012) The BDI 2012 includes the same areas as the BDI 2010. Has 96 items (two per each age level) extracted from the full version of the BDI. It is a screening test that evaluates the child development from 0-8 years old whose use is usually focused on the public sector of education and health. This is an instrument for national and international comparison. It was applied in its full version on the ELPI 2010 which allows to correlate the scores. The objective is to evaluate the fundamental skills development in five areas (personal and social, adaptive, motor, communication and cognitive). Used in ELPI 2012.

Learning and Child Development test (TADI) TADI is a Chilean instrument that allows measuring what children know, and what they do, according to four dimensions of development: language, cognition, motor and socio emotionality, each of which constitutes a separate scale. Therefore, the TADI allows to evaluate learning and development globally, covering the four dimensions, or by each dimension separately: Cognitive that evaluates the concepts of attention, memory, problem solving, mathematical-logical reasoning, knowledge of the world and interest in learning; Motor that is divided in assessing gross motor and fine motor skills of the child; Language that integrates language comprehension skills of listening, speaking, writing and introduction to reading initiation; and Socio emotional that integrates aspects related to independence, self-care, knowledge and self-assessment, recognition and expression of feelings, social interaction, formation of values, self-regulation and close bond. The TADI allows to evaluate children between 3 months and 7 years of age. Used in ELPI 2012.

Measures of child development outcomes - non-cognitive:

Ages and Stages Questionnaires (ASQ: SE) For children 6, 12 and 18 months old, it asks the mothers/caregivers to answer questions as she/he has seen in her child. The alternatives are: Most of the time (0 points), sometimes (5 points) or rarely or never (10 points).Then the mother or caregiver happens to a sheet in which the response alternatives are to remember them better. This test helps identify possible problems in the social and affective development of the child. The item addresses seven behavioural areas: self-regulation, compliance, communication, adaptive functioning, autonomy, affect, and interaction with people. Used in both ELPI 2010 and 2012.

Child Behaviour Checklist (CBCL I and CBCL II) For children between 18-60 months old for CBCL I and above 60 months old for CBCL II, obtain standardised ratings, and descriptive details of children's functioning, as seen by parents/caregivers providing results for three general scales: total problems, internalizing and outsourcing. There are seven syndrome scales designates as emotionally reactive, anxious/depressed, somatic complaints, withdrawn, sleep problems, attention problems, and aggressive behaviour. Used in both ELPI 2010 and 2012.

Measures of maternal abilities: cognitive and non-cognitive

Wechsler Adults Intelligence Scale (WAIS) Language and memory skills (for all ages of mothers or main caregivers). The WAIS measures human intelligence reflected in both verbal (which measures the subject's knowledge of word meaning) and digital distance (ability to recall digits from memory, performance based on the maximum length of a list of digits the subject can recall) abilities. It is based on the belief that intelligence is a global construct, which reflects a variety of measurable skills and that can be considered in the context of the overall personality. The WAIS is also administered as part of a battery test to make inferences about personality and pathology; both through the content of specific answers and patterns of subtest scores. Used in both ELPI 2010 and 2012.

The Big Five Inventory (BFI) For all mothers or main caregivers. The BFI is a self-report inventory designed to measure the Big Five dimensions. The five factors are Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism It contains 44 items and consists of short phrases with relatively accessible vocabulary. Used in both ELPI 2010 and 2012.

Parent Stress Index: Short Form (PSI) For all mothers or main caregivers. The PSI Short Form is a direct derivative of the Parenting Stress Index (PSI) full-length test. All 36 items on the Short Form are contained on the Long Form with identical wording and are written at a 5th-grade reading level, for parents of children 12 years and younger. The PSI/SF yields a Total Stress score from three scales: Parental Distress, Parent-Child Dysfunctional Interaction, and Difficult Child. The PSI/SF was developed at the request of clinicians and researchers who regularly use the full-length PSI and indicated the need for a valid measure administered in less than 10 minutes. It is ideal for clinicians who work in a variety of primary health care settings and have a limited time available to patients, targeting those families most in need of follow-up services. It also is valuable for use in schools and mental health clinics where the parent-child dyad is not the primary focus of the assessment. Used in ELPI 2012.

All Cognitive and Social-Emotional measures have a raw score as well as T scores. As the analysis is within the sample I have internally standardised each measure, taking advantage of the sample size per month, using non-parametric estimation for age and hence I remove the age effect. For this, the first step is to use kernel-weighted local polynomial smoothing methods to regress the raw score on child's age (in months) and recover the mean. The second step is to use again the same estimation method to regress the square of the residuals in the first regression on child's age and recover the variance. Finally, for calculating the z-score, I subtract the mean and divide by the standard deviation for each raw score.

Productivity

The total factor productivity depends on these initial conditions that shifts the production function which are potential characteristics that determines the *initial* level of skills. The covariates that enter into the total factor productivity are child's gender, the birth order, which is fix and do not to change between t and $t+1$ as in the sample only a 5% gave birth to a new child, the birth height and birth weight. Instead of include these parameters as factors in the production function we assume that these only produce a different starting point (the constant term) in the outcome as they are determining initial levels of skills for these different *types* of children and hence produce direct effects on the outcomes and not through the elasticities. The anthropometrics measures were converted to z-scores using the World Health Organization (WHO, 2006).

Measures of parental investment

The Home Observation for Measurement of the Environment (HOME) inventory (Bustos et al., 2001) measures the quality of stimulation and support given to a child in a family atmosphere. The inventory was created by Caldwell et al. (1984, 2003). It consists of 55 items grouped into eight subscales, which records the presence or absence of the trait. This score is obtained from a combination of observation, and a semi-structured interview is conducted in the child's home with the presence of the mother and child. The inventory has eight subscales: Learning Materials, Stimulation of language, Physical environment, Responses from parent to child, Academic stimulation, Modelling and stimulation of social maturity, Diversity of experiences and Acceptance of the child. Nevertheless, in this survey was only collected some items that could be observable during the interview generating new subscales that I will review in the next subsection. Finally, were added three items to observe eating behaviours. The Family Care Indicators (FCI) (Hamadani et al., 2010) is a survey-based indicator of the quality of stimulation of the home environment that help to measure intermediate outcomes and mediators for early childhood development. The FCI were developed to measure home stimulation in large populations and were derived from the Home Observations for Measurement of the Environment (HOME). It has items that can be related to i) parent-child interactions as reading books, singing songs, taking the child outside the home, playing with the child, spending time with the child, etc. and ii) learning materials.

5.3.3 Labour Supply and Parental Investment

A fundamental problem is that most of the research has been based on the effects of parental investment in early childhood development without including in the analysis the labour force participation. This is particularly relevant in Chile and developing countries due to the growth in labour market participation among women with young children. Under this scenario, mothers decrease the amount of time spent with children, but fathers might get more involved in the process of child development and more resources are brought into the family labour income. The question that arises is what is the outcome of children's skills given the fact that more mothers with young children are returning to work by the time their child is less than 6 months old. Hobcraft (1998) found that father's time was especially important for educational outcomes for both boys and girls, and the mother's time was a more consistent predictor for women on outcomes such as the risk of teen parenthood. Figure 5.1 shows the labour market status by child's developmental stages in the sample. It is clear that most of the fathers are employed mothers instead only around 42% of mothers are employed meanwhile around 46%, and 54% are unemployed and inactive respectively.

To study the impact of mother's absence from home or the increase on family labour income on child outcomes the approach should include several steps where parents maximise a constrained model, choosing consumption, saving, hours of work, time investment and monetary investment in their child in a sequential decision problem based on a unitary model. It is important to note that this analysis means to incorporate the time constraint in which it is possible to include the amount of quality time (e.g. reading to the child) the mother and father spend with the child as well as the budget constraint in which is possible to include the monetary investment made by both parents in "didactic materials" to improve child outcomes.

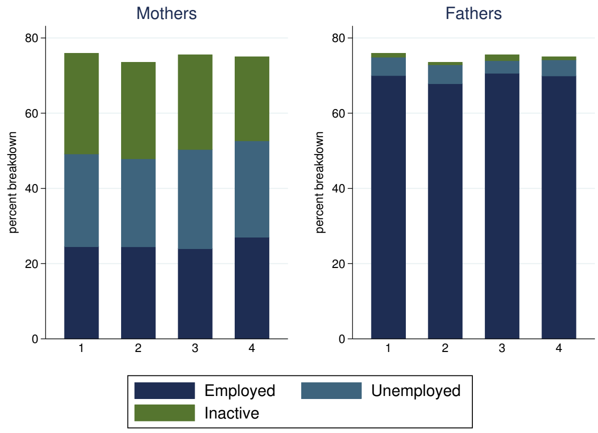


Figure 5.1: Labour Market Status by child's developmental stages

Table 5.1 shows simple associations that I find in the data where the amount of quality time the mother and father spend with the child named as "Parent-Child index" is correlated with child outcomes using cognitive and socioemotional measures at early years. The same is observed regarding monetary investment made by both parents measured using a "Learning materials index".[2] The measures related to Parental-Child activities and Learning Materials at home were constructed with information included in the survey (Appendix 3, Table A3.1 shows the items of each measure). In particular, the activities used to create the Parental-Child factor are read storiesor look atpicture books withchild, tellstories tochild, sing songsto child, take the child to parks, zoo or museums, play with the child and spend time with child talking and/or draw-

[2]Controls used in the analysis are child's weight and length at birth, cognitive and socioemotional mother abilities, socioeconomic status, urban households, parental education, missing parents, number of children and adults in the household and mother's age.

Table 5.1: Child Outcomes and Parental Investment

Measures	Cognitive $(6-24 months)$	Cognitive $(24-58 months)$	Socioemotional $(6-58 months)$
Parent-Child index Mother	0.064***	0.059***	0.049***
	(0.007)	(0.008)	(0.007)
Parent-Child index Father	0.053**	0.057**	0.051**
	(0.029)	(0.023)	(0.020)
Learning materials index	0.009	0.031***	0.015**
	(0.009)	(0.007)	(0.006)
Observations	4807	6639	8963
R-squared	0.105	0.180	0.202

Standard errors (in parentheses), * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

ing pictures, meanwhile the ones to create the Learning materials factor are items that ask if the child has a spacein whichto keep their toys and belongings, one or more games involving muscular activity, toys to push and pull, at least one toy with wheels that can be raised, there are learning materials that are appropriate for age, there are age-appropriate learning equipment and there are literary and musical material.

While it is impossible to account for every influence that contributes to who a child eventually becomes, the literature states that parenting plays an essential role in producing cognitive and non-cognitive skills (Heckman and Masterov, 2004) shaping these abilities through (Cunha and Heckman, 2007) thought it is also crucial to include factors as *initial conditions* or genetics which are related to the period between conception and the child's birth and this period as well as the child environment to understand what determines how a child develops.

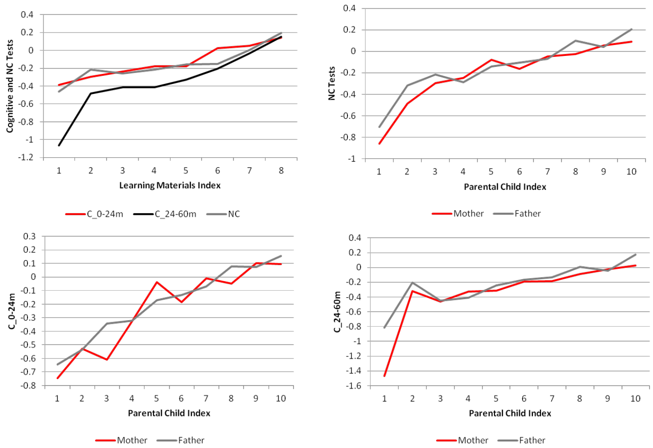


Figure 5.2: Why role of fathers

The fundamental problem is that *parenting* is challenging to measure and hence in most of the cases are not directly observed. Most studies use maternal labour force participation and childcare use in place of maternal time, moreover, the problem is even worst because most studies do not include the role of fathers into account. Figure 5.2 shows that while mother's time is a crucial factor in child development (both skills: cognitive and non-cognitive) father's time is equally critical.

Only by capturing both parents choices can one reasonably expect to answer policy questions that bear upon how households will change their behaviours and the amount of monetary and time investment in response to changes in policies such as paternity leave, childcare facilities or parenting programs.

5.4 Model

There is growing recognition that multiple skills are essential predictors and likely determinants of success in many aspects of life. Although a variety of methods are used to measure these skills, there is no agreement on the best ways to do so. Parental environments and investments at different ages and stages of childhood development determine skills. Recent studies have demonstrated how multiple factors relate in a complex way (Cuhna et al. (2007, 2010)). Understanding the factors affecting the evolution of multiple skills is crucial for the design of effective public policies in developing countries. As pointed out by Cuhna et al. (2010), it is necessary to estimate a multistage technology to capture different development phases in the life cycle of a child. For this, they identified a more general nonlinear technology by extending linear state space and factor analysis to the nonlinear setting. This allows eliminating the assumption that early and late investments are perfect substitutes over the feasible set of inputs. But it is also essential to understand the mechanisms behind parental investment decisions for this reason I develop a model which includes a child production function jointly with a realistic life cycle model that includes parental constraints, fathers choices, unobserved heterogeneity, quality measure of time with the child for both parents, add choices of parental investments not only in a wide range of cognitive skills but also in socioemotional skills, include a measurement system to recover the underlying factors in the estimation of the production functions, and I use a new data for developing countries instead of using PSID and NLSY surveys. For this, I follow a two-step procedure. In the first step, I estimate a measurement model based on a linear dynamic factor model to recover the latent factor, and in the second step, I present a life cycle model of early parental investment in children's human capital that includes preferences, time constraints, budget constraint and the production function for cognitive and socioemotional skills. This model also includes assets, and therefore it is possible to explain how market failures affect the optimal level of parental investments. For example, credit constraints in which there exists liquidity constraints or imperfect markets in which parents cannot borrow against future children's income or own income and therefore cannot leave debt. In both examples, the optimal level of consumption and investment are lower than the optimal. The inclusion of assets also allows analysing precautionary savings due to the multiple sources of risk in the optimal decision which implies over saving and hence the optimal level of consumption and investment are lower than the optimal too.

5.4.1 Step 1: Estimation of the Measurement System

To estimate of the measurement system, I use the procedure proposed by Cuhna et al. (2010). The technologies we will use in the step two include inputs at each t which produce outputs at $t+1$ and children's development is driven by a certain number of latent factors which are reflected in measurements, similarly, investment is also driven home environmental measurements. In the first step, as some issues need to be addressed, such as the presence of measurement error in the data for the inputs in the technology, I estimate the measurement model of the latent factors based on a nonlinear dynamic factor model and exploit cross-equation restrictions (covariance restrictions) proving that I can identify all of them. I consider initially two technologies, one for the production of cognitive skills and one for the production of non-cognitive skills. One innovation in the estimation is to add a new multi-dimensional parental investment, specifically, measured as material resources (monetary investment) and quality time investment.

In the estimation procedure, there are some difficulties that need to be addressed, firstly, we do not observe child's skills, mother's skills and parental investment directly. Indeed, most of the measures used to create factors for cognitive, non-cognitive and investments inputs are measured with error. Parental investment is chosen based on the information from the unobserved variables which is not observed by the econometrician and which includes unobserved shocks or inputs.

For dealing with the first problem, let's use $h_{k,t}$ to explain the procedure, in this case, we can use measurements $y_{k,t,j}$ where each measurement j (e.g. test scores for $k = c$) are additively separable functions in the (log)[3] of the underlaying trait $h_{k,t}$ and has its own informative factor loading, $\alpha_{k,t,j}$ as following:

[3]the use of natural logarithms keeps the latent factor only take positive values.

$$y_{k,t,j} = \mu_{k,t,j} + \alpha_{k,t,j} ln(h_{k,t}) + \varepsilon_{k,t,j}, \quad (5.1)$$

where $\mu_{k,t,j}$ is the mean and the vector $\varepsilon_{k,t,j}$ captures measurement error. For each measurement system, the goal is to recover the latent factor $h_{k,t}$ which is error-free. The identification is achieved setting a factor scale, a factor location, that the errors are uncorrelated across measures and the mean is zero as well as the use of minimum 3 measures foe each latent trait. Using these assumptions it is straightforward to proof the identification for the children's skills, maternal skills and parental investment latent factors.

5.4.2 Step 2: Joint Model

The model describes the sequential decision problem of a household (mother=m, father=f and a child below 7 years old) based on a unitary model. The household is a forward-looking expected utility maximiser endowed with own preferences and makes decisions in each period of a child's developmental stage, where the child's developmental stage is indexed by t. In particular, there are 4 child developmental stages, each of them accounting for 10 months' period, starting at 10 months and ending at 49 months. The child brings utility to their parents through their skills (quality) and at the end of developmental period T. Time is discrete and the horizon is finite (until last development stage). So, the dimension is the developmental stage. Decision period starts at the child's age equal to 10 months old. The last decision period is when the child enters to pre-school at T. Parents make investments in child quality from the first period of the child's life through T, hence, the child brings utility to their parents through the investment in their quality. So, it is a model of preschool investment. The parents make consumption, saving, monetary investment in child, and time allocation decisions. Specifically, in each period, the household choices how much to jointly consume c_t, save a_t, how much time to spend nurturing, monitoring, teaching and caring for the existing child ($qt_{m,t}$ and $qt_{f,t}$), how much to work ($n_{m,t}$ for mother and $n_{f,t}$ for father), and monetary investment on child goods, i_t.

The inclusion of assets in the model is crucial as will allow to analyse failures in the economy as precautionary saving, credit constraints and imperfect markets. For example, if there are multiple sources of risk or negative shocks, then families will prefer to save and the optimal levels of consumption but also parental investment will decrease.

Consumption, saving and monetary investment are continuous decision variables as well as the time allocated to children and the hours of work, though in the discretisation of the hours, I assume three different alternative values for the time spent with the child (low, medium or high) that are matched with the three different alternative values for mother's labour supply and father's labour supply (not work, part-time or full-time employment). Figure 5.3 shows the daily hours of work in the sample for mothers and fathers conditional to be working. Most of the fathers are employed as full-time instead mothers can be categorised into the three groups mentioned before (not work, part-time or full-time employment).

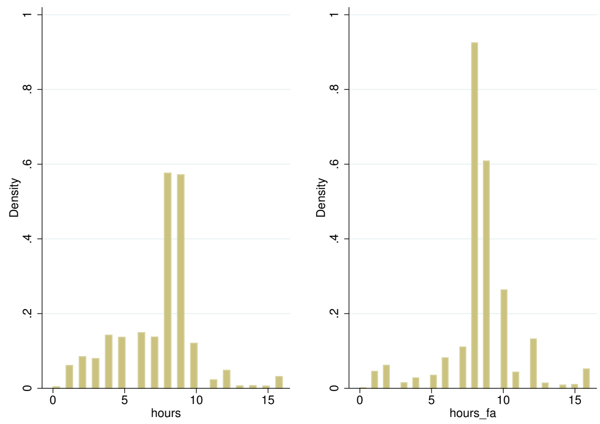


Figure 5.3: Daily hours of work

Parents utility in period t is a function of the level of consumption produced by the household, the level of their child's quality at the terminal period, and the state space. The per-period utility function is (A3) additively separable and has a CRRA form of consumption:

$$U(c_t) = \frac{c_t^{1-\lambda_0}}{1-\lambda_0}, \tag{5.2}$$

The utility function has the usual properties $\partial U/\partial C > 0$, $\partial^2 U/\partial^2 C < 0$. A natural extension of the model is to also incorporate parental labour supply decisions, adding private leisure for parents, $l_{m,t}$ and $l_{f,t}$, in the way of:

$$\begin{aligned} U(c_t) \quad &+ \quad [(\widetilde{\lambda_{1,t}} l_{m,t}^{\phi_0} + (1 - \widetilde{\lambda_{1,t}}) l_{f,t}^{\phi_0})]^{\frac{1}{\phi_0}} \\ &+ \quad \varsigma_t^k, \end{aligned} \tag{5.3}$$

where $\widetilde{\lambda_{1,t}} = \frac{1}{exp(z_t' \lambda_1)}$ which provides observed heterogeneity and ς_t^k unobserved heterogeneity. For the moment, let's continue with (5.2). The parents budget constraint is defined as:

$$c_t + pi_t + a_{t+1} = w_{m,t}(D - qt_{m,t}) + w_{f,t}(D - qt_{f,t}) + y_t + (1 + r)a_t, \tag{5.4}$$

where p is region-level investment prices, r is the interest rate which is equal to $1/(\beta - 1)$, where β is the discount factor, i_t is parental monetary investment, a_t is the level of parental assets, $w_{m,t}$ and $w_{f,t}$ are wages for mother and father respectively and y_t is family not labour income. An extra assumption for the level of savings is that $a_t \geq a$.

In each period, parents are endowed with a fixed amount of time D. They must choose how allocate this time between hours of work $n_{m,t}$ and $n_{f,t}$ respectively and how much time to spend nurturing, monitoring, teaching and caring for the existing child $qt_{m,t}$ and $qt_{f,t}$ respectively:

$$\begin{aligned} n_{m,t} &= D - qt_{m,t} \\ n_{f,t} &= D - qt_{f,t}, \end{aligned} \tag{5.5}$$

As parents are endowed with a fixed amount of time, they choose expenditure in their child, i_t and the quality time with their child, $qt_{m,t}$ and $qt_{f,t}$.

Parents also receive utility through the investment they do in their child's skill levels. Therefore, we also include in the life cycle model two constraints which define the technology of skill formation for cognitive and non-cognitive skills in childhood. The production function is assumed to be the Constant

Elasticity of Substitution (CES) to allow for complementarities of inputs. $h_{c,t}$ denotes cognitive skills of the child at age t, $h_{n,t}$ denotes non-cognitive skills of the child at age t. P_c and P_n represent mother's cognitive and non-cognitive skills respectively which are fix, $qt_{m,t}$, $qt_{f,t}$ and i_t represents now the multi-dimensional parental investment in child skills at age t and A is the total factor productivity. The technology for skill $k = c, nc$ at period t is:

$$\begin{aligned} h_{k,t+1} &= A_k[\gamma_{k,c} h_{c,t}^{\phi_{c,k}} + \gamma_{k,nc} h_{nc,t}^{\phi_{nc,k}} + \gamma_{k,pc} P_c^{\phi_{pc,k}} \\ &+ \gamma_{k,pnc} P_{nc}^{\phi_{pn,k}} + \gamma_{k,qt_m} qt_{m,t}^{\phi_{qt_m,k}} + \gamma_{k,qt_f} qt_{f,t}^{\phi_{qt_f,k}} \\ &+ \gamma_{k,i} i_t^{\phi_{i,k}}]^{\frac{1}{\phi_k}}, \end{aligned} \tag{5.6}$$

where $\gamma_{k,m} \in [0,1]$, $\sum_m \gamma_{k,m} = 1$ for $m \in [c, nc, pc, pn, qt, i]$. The elasticity of substitution is $1/(1-\phi_k)$ and measures the level of substitution in the inputs, where $\phi_k \in (-\infty, 1]$, if $\phi_k \to 1$ then the inputs of the production function become perfect substitutes and if $\phi_k \to -\infty$ they become instead perfect complements. The technology has the regular properties: is monotone increasing in its arguments, twice continuously differentiable, and concave in investment input. This formulation assumes that the formation of skills depend on the initial conditions, A_k, the stock of skills in period t, parental investment at t, i_t and mother's skills, P_k. This formulation provides two key parameters, the self-productivity of skills and a dynamic complementarity. The self-productivity which includes own and cross effects, is the change of the technology when we have a change in the child's skill:

$$\frac{\partial h_{k,t+1}}{\partial h_{k,t}} > 0$$

The dynamic complementarity measures when stocks of skills acquired by the previous period make investment in current period more productive:

$$\frac{\partial h_{k,t+1}^2}{\partial i_t h_{k,t}} > 0$$

The utility parents receive through the investment they do in their child's skill levels is at the end of the final period, for this reason, it is crucial to be explicit about the functional form of the terminal value function, which is defined as:

$$
\begin{aligned}
V_T(h_{c,T}, h_{n,T}, a_T) &= \kappa_c \frac{h_{c,T}^{1-\lambda_0}}{1-\lambda_0} + \kappa_{nc} \frac{h_{nc,T}^{1-\lambda_0}}{1-\lambda_0} \\
&+ \kappa_a[(1 - exp(-a_T)],
\end{aligned} \tag{5.7}
$$

The intuition behind the previous specification is that households make their decisions taking into consideration their expenditure in their final period and the best proxy for this is the level of asset at T. As the value of assets could be negative, we subtracted from 1 so then the terminal value function is not undefined.

The wages are also crucial as these are states of the model in the first three periods. The level of wages affect the budget constraint but they do not affect the level of the terminal value function. The mother and father receive a wage offer of $w_{m,t}$ and $w_{f,t}$ in each period t. Wage offers are not generally observed for non-workers. It is, thus, necessary to specify a wage offer function to carry out estimation. Let wage offers be generated by

$$
w_{z,t} \sim \Gamma(w_{min}, w_t, \epsilon_{z,t}) \quad z \in m, f, \tag{5.8}
$$

Formally, $w_{z,t} = w_{min} + w_{z,t}e^{\epsilon_{z,t}}$ where $\epsilon_{z,t}$ is the income shock. $\epsilon_{z,t}$ is distributed according to a bivariate normal distribution $\epsilon_{z,t} \sim N\left(0, \Sigma\right)$

Before to continue it is useful to distinguish the parents state space, Ω_t, consisting of all of the determinants of parental decisions: $h_{c,t}$, $h_{nc,t}$, P_c, P_n, a_t, t and $\epsilon_{k,t}$, from the part of the state space observable to the researcher, Ω_t^*: $h_{c,t}$, $h_{nc,t}$, P_c, P_n, a_t and t. At any time, parents maximise the present discounted value of lifetime utility, where the solution to the optimisation problem is a set of decision rules between the optimal choice given the state variables at t. The value function can be written as a constrained problem as it follows,

$$
V_t(h_{c,t}, h_{nc,t}, p_c, p_n, a_t; \epsilon_{z,t}) = \max_{c_t, qt_{m,t}, qt_{f,t}, i_t} \left\{ U(c_t) + \beta \int_{\epsilon_{t+1}} V_{t+1}[h_c^{'}, h_{nc}^{'}, p_c, p_n, a_t^{'}; \epsilon_t^{'}] d_\epsilon \right\}
$$

s.t. 5.4, 5.5, 5.6, 5.8, 5.7, $c_t, i_t \geq 0$, $a_t \geq a$, $qt_{m,t} \geq 0$ and $qt_{f,t} \geq 0$. β is the discount factor. The expectation is taken over the distribution of income shocks that differ permanently across households according to unobservables. In the terminal decision period also we have the final period of the parents as it was defined as the child enter to preschool. So, it is in this final period that parents receive their benefits from monetary and time investment through achievement an optimal level of cognitive and non cognitive abilities in their child. The assumption behind is that the parents do not receive benefits meanwhile the child is in the household as they are still investing and hence the effect will be only when the child leaves home achieving positive outcomes.

5.5 Solution and Estimation: work in progress

5.5.1 Solution

The solution of this stochastic optimisation problem needs to be obtained numerically. The model is solved by backward recursion using numerical methods where parents maximise the bellman equation subject to the budget constraint, time constraint for both parents, the technology of skill formation for both skills, the wage formation for both parents and the terminal value function. For do so, I integrate the value function in the next period (V_{t+1}) over income shocks by using standard Gauss-Hermite quadrature methods. The state matrix for the income shocks have 9 combinations as we need to cover the income shock of the mother for each father's income shock. The state variable space contains the variables in Ω_t^*, in specific, the continuos savings represented with $N_a = N$ chebyshev points, the continuos child's skills at developmental age t represented with $N_{h_c} = N$ and $N_{h_{nc}} = N$ chebyshev points, the continuos maternal's skills represented with $N_{p_c} = N$ and $N_{p_n} = N$ chebyshev points and the child-developmental age periods to be 1,2 and 3. The set of control variables are represented by continuos monetary investment, continuos quality time for both parents and the level of consumption. The shocks are integrated by Hermite Quadrature with N_ϵ points for ϵ_{t+1}.

In each period, optimal investments (quality time and monetary), consumption and saving involves the approximation of the expected value function for each possible combination of cognitive and non-cognitive human capital and then over the possible combinations of parental skills, which evolve according to the optimal decisions and the equations governing dynamics (the budget constraint, time constraint, wage equations, and the technology of skill formation for both skills). For the integrated terminal value function, I use Gauss-Hermite quadrature to integrate it: $EV_T(h_{c,T}, h_{nc,T}, a_T, p_c, p_{nc}) = \int_{\epsilon_t} V_t(h_{c,T}, h_{nc,T}, a_T, p_c, p_{nc}, \epsilon_T)$. At every t the input for maximisation is the expected value function,$(EV_{t+1}(h_{c,t+1}, h_{nc,t+1}, a_{t+1}, p_c, p_n))$ which must be approximate. I use linear interpolation, which returns interpolated values of a function of 5 dimensions (state variables) at specific query points using the linear interpolation. The results always pass through the original sampling of the function. For this the coordinates of the sample points are enter into the corresponding function values at each sample point. The outcome of this process contain the coordinates of the chebyshev nodes.

The solution to this problem delivers consistent estimates which are used to solve at every t the Bellman equation at every node and for every shock at t by maximising $V_t(h_c, h_{nc}, \epsilon_t)$ which in terms of grid points is $V_t(N_{h_c}N_{h_{nc}}N_a, N_{p_c}N_{p_n}, N_\epsilon^2, t)$ where:

- N_{h_c} re-scaled chebyshev points of child's cognitive skill grid
- $N_{h_{nc}}$ re-scaled chebyshev points of child's non-cognitive skill grid
- N_a re-scaled chebyshev points of asset grid
- N_{p_c} re-scaled chebyshev points of mother's cognitive skill grid
- N_{p_n} re-scaled chebyshev points of mother's non-cognitive skill grid
- N_ϵ gauss-hermite points for quadrature, one for each shock
- To make value function more manageable I create an index for the the three states that change for every t, $ncna = N_{h_c}N_{h_{nc}}N_a$, an index for the two stages that do not change with t, $np = N_{p_c}N_{p_n}$ and one for both shocks, $ness = N_\epsilon^2$.

Calibration

There are some parameters that are assumed exogenous and hence are not estimated but instead used for the solution of the joint model. The calibration values are based on the fact that the aim of the model is to explain how parents decide their choices more than estimate the correct values of parameters that have been widely study in the literature or that we already have from the observed data. Table 5.2 show the calibration values. For ensuring concavity for the child's cognitive and non-cognitive skills and assets the relative risk aversion is set to 0.5. The intertemporal elasticity of substitution is set to $(1/\lambda_0)$. The discount factor is set to 0.96. The interest rate, r, is derived from the value of the discount factor which is equal to $1/(\beta - 1)$. The relative price of investment to consumption is set to 1 so there is not a relative advantage between both. All the previous values were set using previous literature. For the following I use the observed data, in particular, the minimum level of consumption which is used for the lower bound is set to the minimal wage which is equal in this model for both parents, w_{min} and the wage correlation is set to 0.2. This correlation term is used for the calculation of the Hermite Quadrature's Cholesky Decomposition with N_ϵ^2 points for $\epsilon_{z,t+1}$.

Table 5.2: Calibration Values

	Calibration Value
Relative Risk Aversion λ_0	0.5
Discount Factor (β)	0.96
Relative Price Investment to Consumption (p)	1.0
Minimum Level of Consumption (c_{min})	w_{min}
Wage Correlation (ρ)	0.2

Preliminary Results

Figure A.0.5 in Appendix 7 shows the expected value functions derived for each period t and the terminal period which are concave and monotonic as expected, and where the x-axis is equal to the number of Chevyshev points to evaluate the objective function. As mentioned, this is still a work in progress, and I am currently still solving the model, in particular, the program is running to find a smoother solution using a more efficient approximation method. Once this is done, the derivation of the optimal policy functions $i((h_{c,t}, h_{nc,t}, p_c, p_n, a_t; \epsilon_{z,t})$, $qt_m(h_{c,t}, h_{nc,t}, p_c, p_n, a_t; \epsilon_{z,t})$ and $qt_f(h_{c,t}, h_{nc,t}, p_c, p_n, a_t; \epsilon_{z,t})$ is straightforward and they can be use to simulate data based on random draws of $\epsilon_{z,t}$. Figure A.0.6 in Appendix 7 show preliminary results for the choices made by the family in terms of consumption, c, quality time the mother and father spends with the child qt_m and qt_f as well as the parental monetary investment in child's skills i. There is evidence of a trade-off between quality time and monetary investment which implies lower variability in consumption.

5.5.2 Estimation and Identification

Once the solution is done, the next step is to estimate the model by Simulated Method of Moments (Gourieroux et al (2005), McFadden (1989)). Using this approach the set of data moments is defined by α, and for a given value of the structural parameters (θ), the model is used to simulate the same set of moments $\alpha^s(\theta)$ so that minimising

$$Min_\theta(\alpha - \alpha^s(\theta))'W(\alpha - \alpha^s(\theta))$$

where W is a weighting matrix.

This approach fit simulated data obtained from the model to an auxiliary statistical model that provides a complete statistical description of the data, to be able to identify the behavioural parameters using minimise distance between moments of Chilean data and moments of data simulated using model. The auxiliary statistical models that I plan to use are: multinomial logits for parental labour supply, multinomial logit of transition from employment to non-employment, regression of test scores, logit models for quality parental time, regression of (accepted) wages, among others.

Regarding identification, I use statistical assumptions to identify the within period utility based on the distributional assumptions of shocks required to estimate the model correctly (Magnac and Thesmar, 2002) as well as specify how the initial state space varies across individuals in the sample as this model incorporates unobserved heterogeneity (income shocks).

The number of parameters that need to be estimate is 52 which are presented in groups in Table 5.3. For the estimation of the technology for skill $k = c, nc$ at period t it is needed to estimate six relative productivity parameters for each skill plus the total factor productivity level and the elasticity of substitution between investment and stock of each skill which are fix for each t. It is assumed that the level of the minimum wage is equal for male and females and this is corroborated by the observed data. The preference parameters measure the parental valuation of child's final level of cognitive and non-cognitive skills, κ_c and κ_{nc}, as well as the parental valuation of the level of saving when child enter to the preschool, κ_a. Table 5.3 also shows some example of the auxiliary statistical models that I plan to use for the estimation procedure.

Table 5.3: Structural Parameters

Structural Parameters	Number	Auxiliary Statistical Models
Relative productivity of stock of cognitive skills	$6x3$	Regression of test scores
Relative productivity of stock of non-cognitive skills	$6x3$	Regression of test scores
Elasticity of substitution between investment and stock for both skills	$1x2$	Regression of test scores
Total factor productivity for both skills	$1x2$	Regression of test scores
Minimum wage and wage returns for both parents	$1 + 1x2$	Regression of (accepted) wages
Variance-covariance of income shocks for both parents	$1x2$	Mean and variance of innovation of wage residuals
Preference parameters when child enters preschool	$1x3$	Mean and variance of innovation of wage residuals

5.6 Concluding remarks and further work

Investments in human capital are central to economic growth and for the reduction of inequality. They are particularly relevant for improving the situation of the poor and breaking the intergenerational transmission of poverty. That said, the question of how best to foster human capital formation among the most disadvantaged is not an easy question to answer. There exists increasing evidence that high-quality early childhood investments produce gains in child development that translate into improved long-run outcomes. However, much less is known about the channels through which these early childhood investments operate, and how high-quality investments can be fostered in the family. Moreover, the provision of additional financial resources to low-income families does not automatically translate into better development. This demands a shift in research focus, from understanding the impacts of resources, to understanding the determinants of behaviours.

I develop a dynamic structural model estimated with rich longitudinal data from Chile, in which I integrate a children's human capital model with multiple stages of childhood into a dynamic framework to explain parental investment decisions. Parents maximise a constrained model, choosing consumption, quality time with their child and monetary investments in a sequential decision problem using a unitary model. This way, I explore potential mechanisms: First, the effect of parental preferences when they make decisions in each period of a child's life in terms of his/her developmental outcome measure as cognitive and non-cognitive skills; Second, I analyse the constraints parents face when they are taking their decisions in terms of monetary and quality time with their child; and third, the importance of addressing expectations driving investment choices. An essential contribution to the literature of child development is a three-step procedure used to eliminate the presence of endogeneity (correlation with the unobserved shocks) and measurement error in the data for the inputs in the production function as well as integrating a life cycle model into the analysis. In the first stage, I estimate a measurement model based on a linear dynamic factor model and exploit cross-equation restrictions (covariance restrictions) proving that I can identify all of them. In the second step, I use a dynamic and stochastic structural model that incorporate parental choices based on the overall description of the mechanisms through which parental investment is modified and affects the human capital formation of their children, adding restrictions that involve weaker assumption than those derived from the literature, as well as providing a model for simulating the most effective targeting policies for Early Child Development compensating the most disadvantaged children.

I am currently still in the process of estimating the joint model and finding optimal moments for the estimation of the Simulated Method of Moments as well as calculating the standard errors. The preliminary results suggest that monetary investment decisions are made mostly by mothers, that they also decide no to work mostly in the lower quintiles of the distribution for the first periods but during period three most of then come back to the labour market. Once this model is estimated, I plan to add labour supply decisions in the utility function for both parents as well as incorporate a more complex process for the wages where I could ensure that the moments derived from the data match entirely with the wage distribution generated by the model. I also attempt to incorporate a productivity shock so then unobserved factors are also included in the model which could be associated with for example the access to childcare at each t.

Bibliography

[Achenbach et al. (2000)] Achenbach, T. M., and Rescorla, L. A., "Manual for the ASEBA Preschool Forms Profiles", Burlington: University of Vermont, Research Center for Children, Youth Families.

[Aiyagari et al. (2002)] Aiyagari, S.R, J.Greenwood and A. Seshadri, "Effcient Investment in Children", Journal of Economic Theory, 2002, Vol.102, pp.290-321.

[Apfelbeck et al. (2000)] Apfelbeck, E.M., and Hermosilla, M., "Manual de Administración y Tabulación del Test de WAIS", Santiago.

[Attanasio et al. (2014)] Attanasio, O., Fernández, C., Fitzsimons, E., Grantham-McGregor, S., Meghir, C., and Rubio-Codina, M., "Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial", *BMJ*

[Attanasio et al. (2015)] Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M., "Estimating the Production Function for Human Capital: Results from a Randomized Control Trial in Colombia", *NBER Working Paper* No. 20965 2015.

[Attanasio et al. (2015b)] Attanasio, O., Meghir, C., and Nix, E., "Investments in children and the development of cognition and health in India". Mimeo, Yale University.

[Baydar et al. (1991)] Baydar, N., and Brooks-Gunn, J., "Effects of maternal employment and child-care arrangements on preschoolers cognitive and behavioural outcomes: Evidence from the children of the national longitudinal survey of youth", Developmental Psychology Vol. 27, No. 6, pp 932-45.

[Becker (1964)] Becker, G., "Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education", New York: Columbia University Press.

[Behrman et al. (2006)] Behrman, J.R., Hoddinott, J., Maluccio, J.A., Soler-Hampejsek, E., Behrman, E. L., Martorell, R., Ramirez-Zea, M., and Stein, A. D. (2006), "What Determines Adult Cognitive Skills? Impacts of Pre-Schooling, Schooling and Post-Schooling Experiences in Guatemala", Economics and Demography, University of Pennsylvania.

[Bernal (2008)] Bernal, R., "The effect of maternal employment and child care on children's cognitive development", International Economic Review, Vol. 49, No. 4, pp 1173-1209.

[Bernal et al. (2007)] Bernal, R., and Keane, M., "Child care choices and children?s cognitive achievement: The case of single mothers", Unpublished manuscript, Northwestern University and Yale University.

[Bustos et al. (2001)] Bustos, C., Herrera, M. O., and Mathiesen, M. E., "Calidad del Ambiente del Hogar: Inventario Home Como un Instrumento de Medición", Estudios Pedagógicos, pp 7-22.

[Blau (1999)] Blau, D. M., "The effect of child care characteristics on child development", The Journal of Human Resources, Vol. 34, No. 4, pp 786-822.

[Bronfenbrenner (1979)] Bronfenbrenner, U., "The Ecology of Human Development", Cambridge: Harvard University Press.

[Caldwell et al. (1984, 2003)] Caldwell, B. M., Bradley, R. H., "Home Observation for Measurement of the Environment: Administration Manual". Tempe, AZ: Family Human Dynamics Research Institute, Arizona State University.

[Cárcamo et al. (2015)] Cárcamo, R.A., Vermeer, H.J., van der Veer, R. et al., “Childcare in Mapuche and non-Mapuche families in Chile: The importance of socio-economic inequality”, Journal of Child and Family Studies 24: 2668.

[Carneiro et al. (2003)] Carneiro, P., and Heckman, J. J., “Human Capital Policy”, Inequality in America: What Role for Human Capital Policies?”, Cambridge, MA: MIT Press, pp 77-239.

[Casullo (2000)] Casullo, M., “Big Five Inventory”, Buenos Aires, Argentina: Universidad de Buenos Aires.

[Catherwood (1999)] Catherwood, D., “New views on the young brain: Offerings from developmental psychology to early childhood education. Contemporary Issues in Early Childhood Education”, University College, Worcester, United Kingdom.

[CEDEP] Centro de Estudios de Desarrollo y Estimulación Psicosocial (CEDEP). For organization information see http://www.cedep.info/

[Coddington, et al. (2014)] Coddington, et al. Catherine H. Coddington, , Rashmita S. Mistry, , Alison L. Bailey, “Socioeconomic status and receptive vocabulary development: Replication of the parental investment model with Chilean preschoolers and their families”, Early Childhood Research Quarterly, Volume 29, Issue 4, 4th Quarter 2014, Pages 538-549.

[Contreras, et al. (2015)] Dante Contreras, , Stephanie Gonzz, “Determinants of early child development in Chile: Health, cognitive and demographic factors”, International Journal of Educational Development, Volume 40, January 2015, Pages 217-230.

[Cunha et al. (2013)] Cunha, F., Elo, I., and Culhane, J., “Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation”, *NBER Working Paper* No. 19144 June 2013.

[Cunha et al. (2010)] Cunha, F., and Heckman, J. J. (2010), "Investing in Our Young People", NBER Working Papers 16201, National Bureau of Economic Research, Inc.

[Currie (2001)] Currie, J. (2001), "Early Childhood Education Programs", Journal of Economic Perspectives, Vol. 15, No. 2, pp 213-238.

[Currie et al. (1995)] Currie, J., and Thomas, D., "Does Head Start make a difference?", American Economic Review, Vol. 85, No. 3, pp 341-364.

[Chapman (1988)] Chapman, M., "Constructive evolution: origins and development of Piaget's thought", Cambridge: Cambridge University Press.

[Child (1973)] Child, D., "The Essentials of Factor Analysis", London: Holt, Rinehart Winston.

[CLACYD (2000)] "Estudio Córdoba, Lactancia, Alimentación, Crecimiento y Desarrollo [CLACYD]. (2000), Encuesta de Salud, Nutrición y Desarrollo. Mamás y niel 2000", Ciudad de Córdoba. Córdoba (Argentina): SIMA.

[CLACYD (2002)] "Estudio Córdoba, Lactancia, Alimentación, Crecimiento y Desarrollo [CLACYD] (2002). Iniquidad y desarrollo infantil de 0 a 2 a, Córdoba (Argentina): SIMA.

[Crawford et al. (2010)] Crawford, C., Goodman, A., and Joyce, R., "Explaining the socio-economic gradient in child outcomes: the inter-generational transmission of cognitive skills", Longitudinal and Life Course Studies, Vol. 2, No. 1, pp. 77-93.

[Dahl et al. (2008)] Dahl, G., and Lochner, L., "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit", NBER Working Paper No. 14599.

[Dawson (1994)] Dawson, G., "Development of emotional expression and emotion regulation in infancy", Human behavior and the developing brain, pp 346-79.

[Deaton (2003)] Deaton, A., “Health, inequality and economic development”, Journal of Economic Literature, No. 41, pp 113-58.

[De la Cruz (1998)] De la Cruz, M. V. and González, M., “Inventario de Desarrollo Battelle”, Madrid: Tea Ediciones.

[Diamond et al. (1994)] Diamond, S., Werker, J.F., and Lalonde, C., “Toward understanding commonalities in the development of object search, detour navigation, categorization, and speech perception”, Human behaviour and the developing brain, pp 380-426.

[Duflo (2000)] Duflo, E., “Child Health and Household Resources in South Africa: Evidence from the Old Age Pension Program”, American Economic Review, Vol. 90, No. 2, pp. 393-398.

[Duncan (2003)] Duncan, G. J., “Modeling the impacts of child care quality on children’s preschool cognitive development”, Child Development Vol. 74, No. 5, pp 1454-1475.

[Duncan et al. (1998)] Duncan, G., Young, W., Brooks-Gunn, J., and Smith, J., “How much does childhood poverty affect the life chances of children?”, American Sociological Review, Vol. 63, pp. 406-423.

[Dunn et al. (1986)] Dunn, L. M., Padilla, E., Lugo, D., and Dunn, L. M., “Manual del examinador para el Test de Vocabulario en Imágenes Peabody”, Minneapolis: Pearson.

[ELPI] Early Childhood Longitudinal Survey. For information see http://www.elpi.cl/

[Ermisch et al. (2002)] Ermisch J., Francesconi, M., and Pevalin, D., “Childhood parental behaviour and young people’s outcomes”, ISER Working Paper 2002-12, University of Essex.

[Evans (2004)] Evans, G. W., “The environment of childhood poverty”, American Psychologist, 59, 77?92.

[Filmer et al. (2001)] Filmer, D., and Pritchett, L., “Estimating wealth effect without expenditure data - or tears: An application to educational enrollments in states of India”, Demography, Vol. 38, pp 115-132.

[Gagne (2003)] Gagne, L. G., "Parental work, child-care use, and young children?s cognitive outcomes", Research Paper, Research Data Centres Program. Catalogue no. 89-594 XIE. Ottawa, Ontario, Canada: Statistics Canada.

[Gertler et al. (2004)] Gertler, P. J., and Fernald, L. C., "The Medium Term Impact of Oportunidades on Child Development in Rural Areas", Technical report, University of California at Berkeley.

[Grantham-McGregor et al. (1991)] Grantham-McGregor S., C. Powell, S. Walker and J. Himes., "Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study", *The Lancet* 338(8758): 1-5.

[Hamadani et al. (2010)] Hamadani, J. D., Tofail, F., Hilaly, A., Huda, S. N., Engle, P., Grantham-McGregor, S. M., "Use of family care indicators and their relationship with child development in Bangladesh", Journal of Health, Population, and Nutrition, 28(1), 23?33.

[Haeussler et al. (2003)] Haeussler, I. M., and Marchant, T., "TEPSI Test de Desarrollo Psicomotor 2-5 a, Santiago: Ediciones Universidad Católica de Chile.

[Hair et al. (1987)] Hair, J.; Anderson, R and Tathan, R., "Multivariate Data Analysis readings", Second edition, Macmillan, USA.

[Halpern et al. (1996)] Halpern, R., Barros, F., Horta, B., and Victora, C., "Desenvolvimento neuropsicomotor aos 12 meses de idade em uma coorte de base populacional no Sul do Brasil: diferenciais conforme peso ao nascer e renda familiar." Cad. Saúde Púb, Vol. 12, No. 1, pp 73-78.

[Heckman et al. (2006a)] Heckman, J. J., "Skill formation and the economics of investing in disadvantaged children", Science, Vol. 312, No. 5782, pp 1900-1902.

[Heckman et al. (2007)] Heckman, J. J., and Masterov, D. V., "The Productivity Argument for Investing in Young Children", NBER Working Papers 13016, National Bureau of Economic Research.

[Heckman et al. (2006b)] Heckman, J. J., Stixrud, J., and Urzúa, S., "The Effects of Cognitive and Non-cognitive Abilities on Labor Market Outcomes and Social Behaviour", Journal of Labor Economics Vol. 24, No. 3, pp 411-482.

[Heckman et al. (2006c)] Heckman, J. J., Urzúa, S., and Vytlacil, E. J., "Understanding instrumental variables in models with essential heterogeneity", Review of Economics and Statistics Vol. 88, No. 3, pp 389-432.

[Johnson (1999)] Johnson, M., "A resource sensitive reinterpretation of Lexical Functional Grammar", Journal of Logic, Language and Information, No. 8, pp 45-81.

[JUNJI] Junta Nacional de Jardines Infantiles (JUNJI). For organization information see http://www.junji.cl/portal/index.php

[Kish (1965)] Kish, L., "Surveys Sampling", John Wiley and Sons.

[Kolenikov et al. (2009)] Kolenikov, S., and Angeles, G., "Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer?", Review of Income and Wealth, Vol. 55, No. 1, pp 128-165.

[Lareau (2003)] Lareau, A., "Unequal Childhoods: Race, Class, and Family Life", *Berkeley, CA: University of California Press.*

[Levy et al. (2000)] Levy, D., and Duncan, G., "Using Sibling Samples to Assess the Effect of Childhood Family Income on Completed Schooling", JCPR Working Papers, Northwestern University/University of Chicago Joint Center for Poverty Research.

[Levy et al. (1999)] Levy, P.S., and Lemeshow, S., "Sampling of populations: Methods and applications", Wiley series in probability and statistics. Survey Methodology Section.

[Lewin (1936)] Lewin, K., "A dynamic theory of personality", Nueva York: McGraw-Hill.

[Laitner (1992)] Laitner, J., "Random Earnings Differences, Lifetime Liquidity Constraints and Altruistic Inter-generational Transfers", Journal of Economic Theory, Vol.58, pp.135-170.

[Laitner (1995)] Laitner, J., "Intergenerational and Interhousehold Economic Links", in M. Rosenzweig and Oded Stark eds., Handbook of Population and Family Economics, Amsterdam: North Holland.

[Mayer (1997)] Mayer, S., "What Money Can't Buy: Family Income and Children's Life Chances", Harvard University Press, Cambridge, MA.

[Mincer (1974)] Mincer, J., "Schooling, Experience and Earnings", New York: Columbia University Press.

[McLoyd (1998)] McLoyd, V.C., "Socioeconomic Disadvantage and Child Development", American Psychologist, Vol. 53, No. 2, pp 185-204.

[NSCDC (2010)] National Scientific Council on the Developing Child,"Early Experiences Can Alter Gene Expression and Affect Long-Term Development", Working Paper No. 10. Retrieved from www.developingchild.harvard.edu

[Naudeau et al. (2010)] Naudeau, S., Kataoka, N, Valerio, A., Neuman, M.J., and Elder, L.K., "Investing in Young Children: An Early Childhood Development Guide for Policy Dialogue and Project Preparation", Washington, DC: World Bank.

[Nelson (1995)] Nelson, C.A., "The ontogeny of human memory: A cognitive neuroscience perspective", Developmental Psychology, Vol. 31, pp 723-38.

[Noble et al. (2007)] Noble, K.G., McCandliss, B.D., and Farah, M.J., "Socioeconomic gradients predict individual differences in neurocognitive abilities", Developmental Science, Vol. 10, No. 4, pp 464-80.

[Noboa et al. (2009)] Noboa, G., and Urzúa, S., "Cognitive and non-cognitive development among young children in chile: Effect of participation in public childcare centers", Unpublished manuscript, Northwestern University.

[OECD (2007)] OECD, "Understanding the Brain - The Birth of a Learning Science", Retrieved from http://www.oecd.org/dataoecd/39/53/40554190.pdf

[OECD (2009)] OECD, "Reviews of Labour Market and Social Policies: Chile", OECD Reviews of Labour Market and Social Policies.

[OECD (2011)] OECD, "Society at a Glance 2011 - OECD Social Indicators", Retrieved from www.oecd.org/els/social/indicators/SAG

[Pearson (1901)] Pearson, K., "On Lines and Planes of Closest Fit to Systems of Points in Space", Philosophical Magazine, Vol. 2, No. 6, pp 559-572.

[Paxson et al. (2005)] Paxson, C., and Schady, N., "Cognitive Development among Young Children in Ecuador: The Roles of Wealth, Health and Parenting", Policy Research Working Paper No. 3605. World Bank, Washington, D.C.

[Peairson et al. (2008)] Peairson, S., Austin, A., De Aquino, C., and De Burro, E., "Cognitive Development and Home Environment of Rural Paraguayan Infants and Toddlers Participating in Pastoral del Nin Early Childhood Development Program", Journal of Research in Childhood Education, Vol. 2, No. 4, pp 343-362.

[Piacente et al. (1990)] Piacente, T., Talou, C., and Rodrigo, M., "Piden pan y algo más. Un estudio sobre crecimiento y desarrollo infantil", Madrid: Siglo XXI.

[Putnam (2015)] Putnam, R., "Our Kids: The American Dream in Crisis", *Publisher: Simon Schuster*, 2015.

[Riquelme del Solar (2003)] Riquelme del Solar, G., "Test de habilidades básicas para la iniciación al cálculo TIC", Revista Enfoques Educacionales, Vol. 5, No. 1, pp 137-156.

[Rodriguez et al. (2008)] Rodriguez, S., Arancibia, V., and Undurraga, C., "Escala de Evaluación cognitivas de 0-24 Meses", Santiago: Cronopios.

[Schellenberg et al. (2003)] Schellenberg, J. A., Victora, C.G., Mushi, A., et al., “Inequities among the very poor: health care for children in southern Tanzania”, The Lancet, Vol. 361, No. 9357, pp 561-6.

[Shea (2000)] Shea, J., “Does parents’ money matter?”, Journal of Public Economics, Vol. 77, No. 2, pp. 155-184.

[Shonkoff et al. (2000)] Shonkoff, J.P., and Phillips, D.A., “From Neurons to Neighbourhoods: The Science of Early Childhood Development”, National Academy Press, Washington, DC.

[Squires et al. (2000)] Squires, J., Bricker, D., and Twombly, E., “The ASQ:SE User’s Guide”, Baltimore: Paul H. Brookes Publishing.

[TSECD (2007)] The Science of Early Childhood Development, National Scientific Council on the Developing Child. Retrieved from http://www.developingchild.net

[Thompson (2001)] Thompson, R. A., “Development in the first years of life”, The Future of Children, Vol. 11, No. 1, pp 20-33.

[UNDP (2008)] UNDP, “Human Development Report 2008/2009”, New York, NY: United Nations Development Programme.

[UNESCO (2010)] UNESCO, “Education for All Global Monitoring Report: Reaching the marginalized”, Paris, France: UNESCO

[UNESCO] Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO). For organization information see http://www.unesco.org/new/es/unesco/

[UNICEF (2004)] UNICEF, “La Equidad se Juega en la Primera Infancia”, Working Paper No. 4.

[Van der Gaag (2005)] Van der Gaag, J, “Organization of American States”.

[Vega (2011)] Vega, J., “Steps toward the health equity agenda in Chile. Background paper for World Health Organization?s World Conference on the

Social Determinants of Health", Retrieved from the World Health Organization?s website: www.who.int/sdhconference/.../draft background paper25 chile.pdf

[Walker et al. (2007)] Walker, S. P., Wachs, T. D., Gardner, J. M., Lozoff, B., Wasserman, G. A., Pol- litt, E., et al, "Child development: Risk factors for adverse outcomes in developing countries", Lancet, 369, 145?157.

[Wechsler (1974)] Wechsler, D., "Wechsler intelligence scale for children-Revised edition", New York: Psychological Corporation.

[WHO (2006)] World Health Organization, "Introducing the New WHO Child Growth Standards", Report of a Regional Workshop Bangkok, Thailand, 5-7 June 2006. Retrieved from http://203.90.70.117/PDS˙DOCS/B0390.pdf

Appendix A

Appendix

Appendix 1: Wealth Index

The Overall Kaiser-Meyer-Olkin is a measure of sampling adequacy given by the following criterion:

Table A.1: Kaiser-Meyer-Olkin measure

0.00 to 0.49	unacceptable
0.50 to 0.59	miserable
0.60 to 0.69	mediocre
0.70 to 0.79	middling
0.80 to 0.89	meritorious
0.90 to 1.00	marvellous

Table A1 1: Descriptive Statistics Wealth index and Polychoric Principal Component Analysis

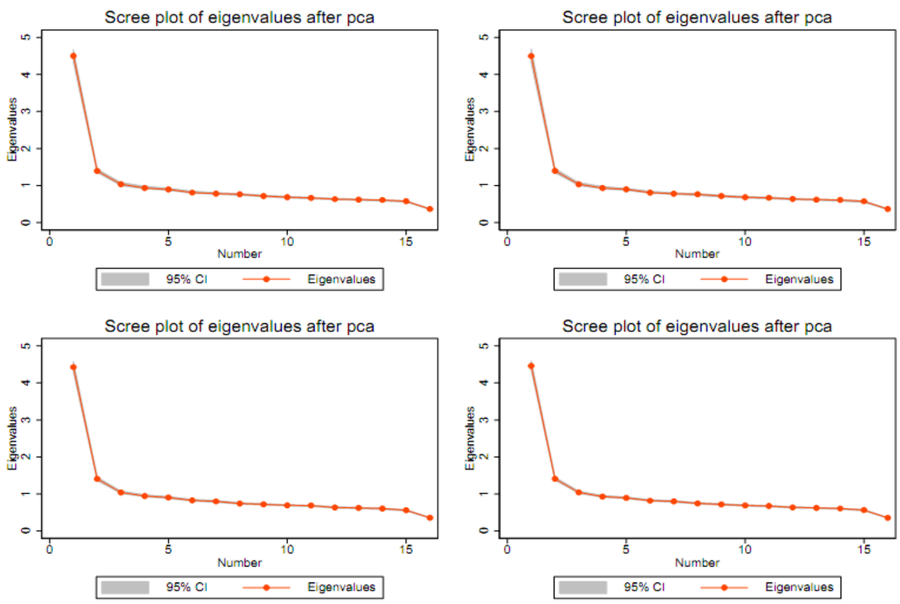
	WEALTH												
Variables	Overall	EEDP			BDI			TVIP			TEPSI		
	smc*	Mean	St. Dev.	Component	Mean	St. Dev.	Component	Mean	St. Dev.	Component	Mean	St. Dev.	Component
Refrigerator	0.1842	2.71	1.19	-0.511	2.71	1.18	-0.513	2.73	1.19	-0.499	2.72	1.20	-0.504
		0.91	0.29	0.050	0.91	0.28	0.050	0.92	0.27	0.042	0.92	0.27	0.045
Washing machine	0.2684			-0.355			-0.355			-0.373			-0.373
		0.74	0.44	0.122	0.74	0.44	0.123	0.76	0.43	0.117	0.75	0.43	0.121
VCR or DVD	0.1674			-0.255			-0.255			-0.264			-0.262
		0.68	0.47	0.119	0.68	0.47	0.120	0.70	0.46	0.111	0.70	0.46	0.113
Microwave	0.2803			-0.234			-0.234			-0.227			-0.230
		0.52	0.50	0.213	0.52	0.50	0.214	0.52	0.50	0.206	0.53	0.50	0.206
Calefont or thermo-electric	0.3209	0.62	0.49	-0.299	0.62	0.48	-0.302	0.63	0.48	-0.306	0.62	0.49	-0.301
				0.183			0.183			0.183			0.184
Video camera	0.2416	0.14	0.35	-0.076	0.14	0.35	-0.076	0.13	0.34	-0.072	0.13	0.34	-0.071
				0.453			0.450			0.467			0.459
Cell phone contract	0.2692	0.27	0.45	-0.120	0.27	0.44	-0.120	0.26	0.44	-0.120	0.26	0.44	-0.119
				0.320			0.321			0.338			0.334
Internet connection	0.4770	0.28	0.45	-0.150	0.28	0.45	-0.150	0.29	0.46	-0.159	0.29	0.45	-0.157
				0.379			0.380			0.385			0.385
Desktop computer	0.3519	0.39	0.49	-0.166	0.39	0.49	-0.166	0.40	0.49	-0.174	0.39	0.49	-0.171
				0.261			0.261			0.260			0.263
Notebook, laptop	0.3192	0.20	0.40	-0.097	0.20	0.40	-0.096	0.20	0.40	-0.098	0.20	0.40	-0.100
				0.377			0.378			0.388			0.386
Cable or satellite TV	0.2170	0.43	0.50	-0.166	0.43	0.50	-0.167	0.45	0.50	-0.175	0.44	0.50	-0.173
				0.219			0.218			0.218			0.219
Predominant material in housing	0.1238	4.48	0.95	0.000	4.48	0.94	0.000	4.46	0.98	0.000	4.46	0.98	0.000
Ground floor				-0.670			-0.669			-0.640			-0.629
Wood or plastic				-0.418			-0.417			-0.389			-0.385
Table or parquet				-0.260			-0.259			-0.240			-0.238
No coated radier				-0.168			-0.168			-0.157			-0.155
Coated radier				0.100			0.099			0.097			0.096
Rent	0.2186	100011	90447	0.255	99940	90715	0.255	101060	91576	0.226	101518	94755	0.228
Overall Kaiser-Meyer-Olkin			0.8928			0.8925			0.888			0.8893	
Variation explained			49.7%			49.7%			48.3%			48.6%	

Note: Polychoric PCA with one principal component

* Squared multiple correlations of variables with all other variables

Figure A1 1: Scree plot criterion (optimal component number)

EEDP BDI



TVIP TEPSI

Table A1 2: Correlation Polychoric PCA, FA and PCA for Wealth index and all cognitive measures

EEDP

	fi	fm1	pc	f
fi	1			
fm1	1.000	1		
pc	0.996	0.995	1	
f	0.997	0.997	0.996	1

TVIP

	fi	fm1	pc	f
fi	1			
fm1	0.999	1		
pc	0.995	0.994	1	
f	0.997	0.995	0.997	1

BDI

	fi	fm1	pc	f
fi	1			
fm1	1.000	1		
pc	0.996	0.995	1	
f	0.997	0.997	0.996	1

PPVT

	fi	fm1	pc	f
fi	1			
fm1	0.999	1		
pc	0.995	0.994	1	
f	0.997	0.996	0.997	1

Note: f (Polychoric PCA); fi (FA with iterative extraction method); fm (FA with ml extraction method); pc (PCA).

Appendix 2: Socioeconomic Status Index

Table A2 1: Descriptive Statistics SES index and Principal Component Analysis

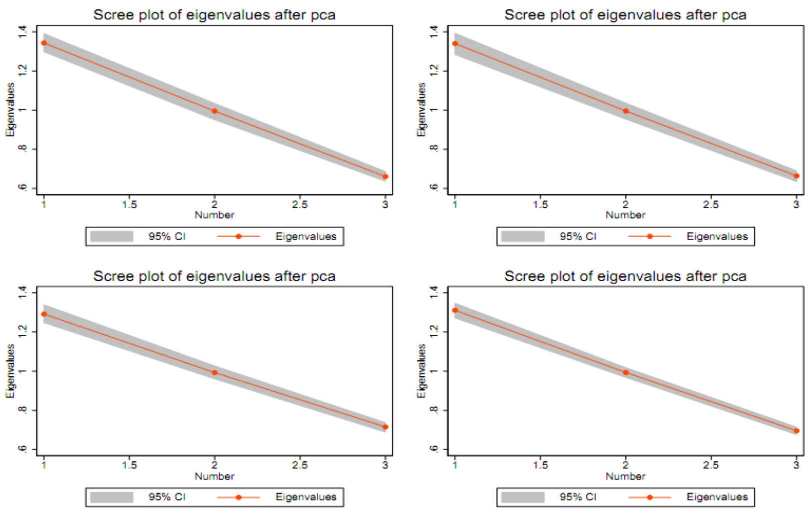
		Socioeconomic Status											
Variables	Overall	EEDP			BDI			TVIP			TEPSI		
	smc*	Mean	St. Dev.	Component	Mean	St. Dev.	Component	Mean	St. Dev.	Component	Mean	St. Dev.	Component
Parental years of schooling	0.115	9.68	4.60	0.70	9.70	4.59	0.70	9.89	4.47	0.70	9.88	4.49	0.70
Parental occupational position	0.086	3.14	0.46	0.12	3.13	0.46	0.12	3.14	0.47	0.15	3.14	0.47	0.14
Per capita income	0.115	137.51	161.30	0.70	137.29	161.65	0.70	139.34	176.80	0.70	139.34	171.83	0.70
Overall Kaiser-Meyer-Olkin			0.5022			0.5019			0.5039			0.5037	
Variation explained			44.8%			44.7%			43.1%			43.7%	

Note: PCA with one principal component

* Squared multiple correlations of variables with all other variables

Figure A2 1: Scree plot criterion (optimal component number)

EEDP BDI



TVIP TEPSI

Table A2 2: Correlation FA and PCA for SES index and all cognitive measures

EEDP	fi	fm1	f
fi	1		
fm1	0.999	1	
pca	0.941	0.923	1

TVIP	fi	fm1	f
fi	1		
fm1	1.000	1	
pca	0.995	0.993	1

BDI	fi	fm1	f
fi	1		
fm1	0.999	1	
pca	0.897	0.871	1

PPVT	fi	fm1	f
fi	1		
fm1	1.000	1	
pca	0.997	0.996	1

Note: fi (FA with iterative extraction method); fm (FA with ml extraction method); pc (PCA).

Appendix 3: HOME

Table A3 1: Descriptive Statistics HOME and FCI for wave 2010

HOME*															
Subscale and Items	EEDP			BDI			TVIP			TEPSI			CBCL1		
1. Emotional and verbal responsivity	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.
Spontaneously vocalises to child at least twice	4955	0.97	0.18	4807	0.97	0.18	6639	0.95	0.21	8963	0.95	0.21	11184	0.96	0.21
Speech is distint, clear and audible	4955	0.96	0.19	4807	0.96	0.19	6639	0.96	0.18	8963	0.96	0.19	11184	0.96	0.19
Initiates verbal interchanges with interviewer	4955	0.91	0.28	4807	0.91	0.28	6639	0.91	0.28	8963	0.91	0.29	11184	0.91	0.29
Expresses ideas easily, speaks fluently	4955	0.90	0.30	4807	0.90	0.30	6639	0.91	0.29	8963	0.90	0.30	11184	0.90	0.30
Spontaneously praises child at least twice	4955	0.84	0.37	4807	0.84	0.37	6639	0.76	0.43	8963	0.75	0.43	11184	0.77	0.42
Caresses or kisses child at least once	4956	0.86	0.34	4808	0.86	0.34	6640	0.72	0.45	8964	0.73	0.45	11185	0.75	0.43
Response positively to praise of child	4956	0.91	0.29	4808	0.91	0.29	6640	0.85	0.36	8964	0.84	0.37	11185	0.86	0.35
2. Parental involvement															
Keeps within his visual range and look often	4956	0.90	0.29	4808	0.91	0.29	6640	0.81	0.40	8964	0.81	0.39	11185	0.82	0.38
Talks to the child while answering the survey	4955	0.86	0.34	4807	0.86	0.34	6639	0.74	0.44	8963	0.76	0.43	11184	0.78	0.42
Consciously encourages development advance	4955	0.75	0.43	4807	0.76	0.43	6639	0.69	0.46	8963	0.69	0.46	11184	0.70	0.46
Invests maturing toys with educational value	4956	0.53	0.50	4808	0.53	0.50	6640	0.45	0.50	8964	0.44	0.50	11185	0.46	0.50
Mothers structured period of play the child	4956	0.36	0.48	4808	0.36	0.48	6640	0.35	0.48	8964	0.34	0.47	11185	0.35	0.48
Provides toys that challenge child	4956	0.45	0.50	4808	0.45	0.50	6640	0.40	0.49	8964	0.39	0.49	11185	0.40	0.49
FCI: Parent-Child Activities**															
Items	EEDP			BDI			TVIP			TEPSI			CBCL1		
	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.
Read stories or look at picture books with child	4956	0.69	0.46	4808	0.69	0.46	6641	0.79	0.40	8966	0.78	0.41	11193	0.77	0.42
Tell stories to child	4956	0.68	0.47	4808	0.68	0.47	6641	0.82	0.38	8966	0.81	0.40	11193	0.79	0.41
Sing songs with child	4956	0.98	0.15	4808	0.98	0.15	6641	0.96	0.19	8966	0.96	0.19	11193	0.97	0.18
Take the child to parks, zoo or museums	4956	0.79	0.40	4808	0.79	0.40	6641	0.82	0.39	8966	0.81	0.39	11193	0.81	0.39
Play with the child	4956	0.99	0.08	4808	0.99	0.08	6641	0.99	0.10	8966	0.99	0.10	11193	0.99	0.10
Mother/caregiver reads stories at least 3 times a week	4956	0.58	0.49	4808	0.58	0.49	6641	0.67	0.47	8966	0.66	0.48	11070	0.64	0.48
Spend time with child talking and/or drawing pictures	4956	0.93	0.25	4808	0.93	0.25	6641	0.97	0.16	8966	0.97	0.17	11193	0.97	0.18
FCI: Learning Materials***															
Items	EEDP			BDI			TVIP			TEPSI			CBCL1		
	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.
Has a space in which to keep their toys and belongings	4956	0.88	0.32	4808	0.88	0.32	6641	0.90	0.30	8966	0.89	0.31	11168	0.89	0.31
Has one or more games involving muscular activity	4956	0.77	0.42	4808	0.77	0.42	6641	0.82	0.39	8966	0.82	0.39	11157	0.82	0.39
Has to push and pull toys	4956	0.80	0.40	4808	0.80	0.40	6641	0.86	0.35	8966	0.87	0.34	11171	0.87	0.34
Has at least one toy with wheels that can be raised	4956	0.72	0.45	4808	0.72	0.45	6641	0.88	0.32	8966	0.88	0.33	11169	0.86	0.34
There are learning materials that are appropriate for age	4956	0.89	0.31	4808	0.89	0.31	6641	0.92	0.27	8966	0.91	0.28	11170	0.91	0.29
There are age-appropriate learning equipment	4956	0.78	0.42	4808	0.78	0.42	6641	0.68	0.47	8966	0.69	0.46	11157	0.70	0.46
There are literary and musical material	4956	0.76	0.42	4808	0.76	0.42	6641	0.81	0.40	8966	0.79	0.40	11154	0.79	0.41
Child has 3 or more books his/her property	4956	0.62	0.49	4808	0.62	0.49	6641	0.78	0.42	8966	0.76	0.43	11127	0.74	0.44

* Actions taken by mother/caregiver during the interview

** Actions taken by mother/caregiver, father or other relative

***Actions taken by child

Table A3 2: Descriptive Statistics HOME and FCI for wave 2012

HOME*												
Subscale and Items	TADI			BDI			TVIP			CBCL1		
1. Emotional and verbal responsivity	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.
Converses with child two or more times during visit	10537	0.92	0.26	10658	0.92	0.26	10399	0.92	0.26	9034	0.93	0.26
Answers child's questions or requests verbally	10544	0.94	0.25	10665	0.94	0.25	10406	0.94	0.25	9034	0.94	0.24
Usually responds verbally to child's speech	10544	0.95	0.23	10665	0.95	0.23	10406	0.95	0.23	9034	0.95	0.22
Praises child's qualities twice during visit	10544	0.70	0.46	10665	0.70	0.46	10406	0.70	0.46	9034	0.70	0.46
Caresses, kisses, or cuddles child during visit	10544	0.62	0.49	10665	0.62	0.49	10406	0.62	0.49	9034	0.64	0.48
Helps child demonstrate some achievement during visit	10544	0.68	0.46	10665	0.68	0.47	10406	0.68	0.46	9034	0.69	0.46
2. Learning Materials												
Two or more toys which teach colors, sizes and shapes are available	10544	0.55	0.50	10665	0.55	0.50	10406	0.56	0.50	9034	0.55	0.50
Three or more puzzles are available	10544	0.42	0.49	10665	0.42	0.49	10406	0.42	0.49	9034	0.42	0.49
Record player or tape recorder and five or more children's records or tapes are available	10544	0.73	0.44	10665	0.73	0.44	10406	0.73	0.44	9034	0.73	0.44
Two or more toys or games permitting free expression are available	10544	0.62	0.49	10665	0.62	0.49	10406	0.62	0.49	9034	0.62	0.49
Two or more toys or games which help teach numbers are available	10544	0.41	0.49	10665	0.41	0.49	10406	0.42	0.49	9034	0.41	0.49
Ten or more books are available	10544	0.31	0.46	10665	0.31	0.46	10406	0.31	0.46	9034	0.30	0.46
FCI: Parent-Child Activities**												
Items	TADI			BDI			TVIP			CBCL1		
	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.
Read stories or look at picture books with child	10518	0.81	0.39	10560	0.81	0.39	10370	0.82	0.39	8999	0.81	0.39
Tell stories to child	10528	0.83	0.37	10569	0.83	0.37	10381	0.83	0.37	9013	0.84	0.37
Sing songs with child	10651	0.90	0.30	10689	0.90	0.30	10502	0.90	0.29	9118	0.91	0.28
Take the child to park	10274	0.72	0.45	10307	0.71	0.45	10124	0.72	0.45	8777	0.72	0.45
Take the child to library, zoo or museums	9653	0.35	0.48	9688	0.34	0.48	9517	0.35	0.48	8256	0.35	0.48
Spend time with child talking and/or drawing pictures	10732	0.97	0.18	10776	0.97	0.18	10583	0.97	0.18	9184	0.97	0.17
Teach animal sounds to the child	10301	0.90	0.30	10337	0.90	0.30	10148	0.90	0.30	8903	0.91	0.29
Teach colours to the child	10241	0.91	0.29	10278	0.91	0.29	10090	0.91	0.29	8857	0.92	0.28
Teach numbers to the child	10504	0.94	0.24	10542	0.94	0.24	10349	0.94	0.24	9041	0.94	0.24
Teach letters to the child	10484	0.92	0.27	10525	0.92	0.27	10331	0.92	0.27	9019	0.92	0.27

* Actions taken by mother/caregiver during the interview

** Actions taken by mother/caregiver, father or other relative

Table A3 3: Factor Analysis HOME for all cognitive measures in both waves

HOME_2012

Items	Overall	TADI		BDI		TVIP		CBCL1	
	smc*	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
Converses with child two or more times during visit	0.862		0.652		0.651		0.655		0.632
Answers child's questions or requests verbally	0.736		0.774		0.774		0.772		0.771
Usually responds verbally to child's speech	0.748		0.757		0.757		0.755		0.749
Praises child's qualities twice during visit	0.833		0.452		0.451		0.455		0.449
Caresses, kisses, or cuddles child during visit	0.853		0.379		0.380		0.381		0.374
Helps child demonstrate some achievement during visit	0.820		0.413		0.412		0.416		0.411
Two or more toys which teach colors, sizes and shapes are available	0.879	0.693		0.693		0.694		0.693	
Three or more puzzles are available	0.872	0.717		0.718		0.718		0.721	
Record player or tape recorder and five or more children's records or tapes	0.887	0.433		0.434		0.430		0.437	
Two or more toys or games permitting free expression are available	0.882	0.587		0.588		0.586		0.583	
Two or more toys or games which help teach numbers are available	0.864	0.744		0.743		0.744		0.738	
Ten or more books are available	0.883	0.626		0.626		0.627		0.618	
Overall Kaiser-Meyer-Olkin		0.8054		0.8057		0.8053		0.8016	
Variation explained		56.3%		56.5%		56.0%		58.7%	

HOME_2010

Items	Overall	EEDP		BDI		TVIP		TEPSI		CBCL1	
	smc*	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
Spontaneously vocalises to child at least twice	0.215		0.392		0.393		0.312		0.346		0.348
Speech is distint, clear and audible	0.346		0.635		0.635		0.550		0.587		0.592
Initiates verbal interchanges with interviewer	0.324		0.547		0.548		0.487		0.482		0.494
Expresses ideas easily, speaks fluently	0.392		0.692		0.690		0.737		0.722		0.724
Spontaneously praises child at least twice	0.395		0.250		0.251		0.228		0.229		0.229
Caresses or kisses child at least once	0.273		0.127		0.129		0.136		0.141		0.138
Response positively to praise of child	0.339		0.117		0.116		0.118		0.128		0.124
Keeps within his visual range and look often	0.158	0.248		0.240		0.384		0.369		0.364	
Talks to the child while answering the survey	0.242	0.349		0.343		0.475		0.460		0.459	
Consciously encourages development advance	0.360	0.445		0.468		0.547		0.527		0.520	
Mothers structured period of play the child	0.334	0.263		0.264		0.340		0.332		0.316	
Overall Kaiser-Meyer-Olkin	0.8329			0.8329		0.8302		0.8309		0.832	
Variation explained	77.4%			77.5%		78.0%		78.0%		78.0%	

Note: Factor Analysis with maximum likelihood extraction method and varimax rotation excluded 3 factors

* Squared multiple correlations of variables with all other variables

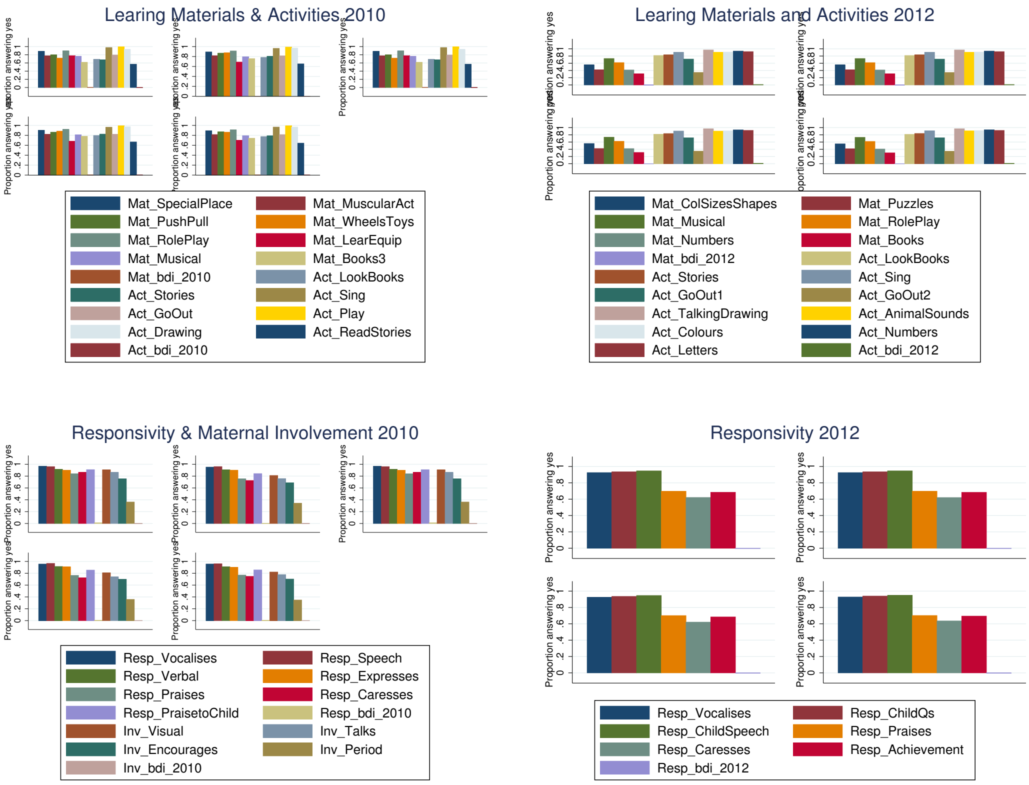


Figure A.0.1: Home Assessments: Proportion answering yes

Table A.2: IRT analysis Lear_Mat_2010

	(1) EEDP	(2) TEPSI	(3) BDI	(4) TVIP	(5) CBCL1
Mat_SpecialPlace					
αi	1.511***	1.643***	1.509***	1.592***	1.672***
	(0.086)	(0.068)	(0.086)	(0.078)	(0.062)
βi	2.733***	2.974***	2.729***	3.041***	2.998***
	(0.087)	(0.071)	(0.086)	(0.084)	(0.065)
Mat_MuscularAct					
αi	1.888***	1.770***	1.885***	1.727***	1.796***
	(0.095)	(0.068)	(0.095)	(0.077)	(0.062)
βi	1.915***	2.231***	1.913***	2.244***	2.228***
	(0.074)	(0.057)	(0.073)	(0.066)	(0.052)
Mat_PushPull					
αi	2.198***	2.065***	2.197***	2.021***	2.121***
	(0.121)	(0.084)	(0.121)	(0.095)	(0.078)
βi	2.333***	3.016***	2.331***	2.920***	3.073***
	(0.097)	(0.083)	(0.097)	(0.092)	(0.077)
Mat_WheelsToys					
αi	1.604***	1.915***	1.605***	1.885***	1.854***
	(0.079)	(0.077)	(0.079)	(0.089)	(0.066)
βi	1.355***	2.996***	1.354***	3.068***	2.774***
	(0.055)	(0.078)	(0.055)	(0.092)	(0.064)
Mat_RolePlay					
αi	1.210***	1.653***	1.210***	1.623***	1.617***
	(0.076)	(0.071)	(0.076)	(0.084)	(0.063)
βi	2.633***	3.270***	2.633***	3.354***	3.193***
	(0.077)	(0.080)	(0.077)	(0.095)	(0.069)
Mat_LearEquip					
αi	1.176***	1.428***	1.177***	1.407***	1.399***
	(0.064)	(0.052)	(0.064)	(0.060)	(0.046)
βi	1.575***	1.088***	1.575***	1.040***	1.145***
	(0.051)	(0.035)	(0.051)	(0.040)	(0.032)
Mat_Musical					
αi	1.624***	1.778***	1.633***	1.782***	1.826***
	(0.083)	(0.067)	(0.083)	(0.079)	(0.062)
βi	1.696***	2.043***	1.701***	2.178***	2.046***
	(0.063)	(0.054)	(0.063)	(0.066)	(0.050)
Mat_Books3					
αi	1.106***	1.243***	1.110***	1.264***	1.238***
	(0.056)	(0.048)	(0.056)	(0.057)	(0.042)
βi	0.573***	1.468***	0.577***	1.632***	1.359***
	(0.038)	(0.037)	(0.038)	(0.045)	(0.032)
var(Theta)					
βi	1	1	1	1	1
	(.)	(.)	(.)	(.)	(.)
N	4864	9156	4865	6844	11181

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: IRT analysis Act_2010

	(1) EEDP	(2) TEPSI	(3) BDI	(4) TVIP	(5) CBCL1
Act_LookBooks					
αi	2.442***	2.509***	2.441***	2.422***	2.436***
	(0.137)	(0.115)	(0.137)	(0.128)	(0.097)
βi	1.471***	2.384***	1.469***	2.469***	2.258***
	(0.078)	(0.085)	(0.078)	(0.099)	(0.070)
Act_Stories					
αi	2.906***	2.793***	2.897***	2.803***	2.829***
	(0.194)	(0.138)	(0.192)	(0.164)	(0.126)
βi	1.566***	2.839***	1.558***	3.070***	2.667***
	(0.097)	(0.110)	(0.097)	(0.138)	(0.094)
Act_Sing					
αi	1.667***	1.702***	1.666***	1.629***	1.688***
	(0.155)	(0.098)	(0.155)	(0.110)	(0.090)
βi	4.957***	4.458***	4.957***	4.360***	4.524***
	(0.224)	(0.132)	(0.224)	(0.146)	(0.122)
Act_GoOut					
αi	0.639***	0.659***	0.640***	0.636***	0.679***
	(0.049)	(0.039)	(0.049)	(0.045)	(0.035)
βi	1.464***	1.608***	1.466***	1.629***	1.614***
	(0.041)	(0.032)	(0.041)	(0.037)	(0.029)
Act_Play					
αi	1.868***	2.255***	1.868***	2.208***	2.119***
	(0.249)	(0.189)	(0.249)	(0.219)	(0.161)
βi	6.669***	6.850***	6.670***	6.792***	6.644***
	(0.447)	(0.339)	(0.447)	(0.391)	(0.285)
Act_Drawing					
αi	1.209***	1.694***	1.209***	1.613***	1.664***
	(0.087)	(0.103)	(0.087)	(0.120)	(0.089)
βi	3.154***	4.681***	3.154***	4.710***	4.511***
	(0.095)	(0.144)	(0.095)	(0.167)	(0.121)
Act_ReadStories					
αi	1.929***	1.677***	1.933***	1.721***	1.757***
	(0.094)	(0.063)	(0.094)	(0.077)	(0.059)
βi	0.440***	0.925***	0.444***	1.035***	0.880***
	(0.048)	(0.037)	(0.048)	(0.045)	(0.034)
var(Theta)					
βi	1	1	1	1	1
	(.)	(.)	(.)	(.)	(.)
N	4868	9167	4869	6852	11193

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: IRT analysis Resp_2010

	(1) EEDP	(2) TEPSI	(3) BDI	(4) TVIP	(5) CBCL1
Resp_Vocalises					
αi	2.795***	2.475***	2.789***	2.542***	2.553***
	(0.225)	(0.126)	(0.224)	(0.155)	(0.121)
βi	5.963***	5.002***	5.957***	5.181***	5.223***
	(0.333)	(0.170)	(0.332)	(0.213)	(0.167)
Resp_Speech					
αi	2.329***	1.720***	2.324***	1.640***	1.736***
	(0.178)	(0.095)	(0.178)	(0.112)	(0.087)
βi	5.035***	4.336***	5.029***	4.414***	4.344***
	(0.244)	(0.124)	(0.243)	(0.147)	(0.113)
Resp_Verbal					
αi	2.421***	1.992***	2.410***	1.981***	1.993***
	(0.151)	(0.083)	(0.151)	(0.098)	(0.076)
βi	4.007***	3.477***	3.997***	3.575***	3.527***
	(0.172)	(0.092)	(0.171)	(0.110)	(0.085)
Resp_Expresses					
αi	2.144***	1.788***	2.140***	1.706***	1.789***
	(0.133)	(0.075)	(0.132)	(0.086)	(0.069)
βi	3.493***	3.185***	3.484***	3.222***	3.189***
	(0.141)	(0.080)	(0.141)	(0.093)	(0.073)
Resp_Praises					
αi	3.368***	4.253***	3.363***	4.434***	4.218***
	(0.248)	(0.278)	(0.247)	(0.350)	(0.245)
βi	3.727***	3.134***	3.708***	3.409***	3.344***
	(0.215)	(0.173)	(0.214)	(0.226)	(0.162)
Resp_Caresses					
αi	1.744***	1.909***	1.742***	1.936***	1.937***
	(0.097)	(0.065)	(0.097)	(0.077)	(0.061)
βi	2.662***	1.543***	2.659***	1.534***	1.720***
	(0.091)	(0.047)	(0.091)	(0.055)	(0.046)
Resp_PraisetoChild					
αi	2.643***	2.591***	2.639***	2.537***	2.688***
	(0.166)	(0.098)	(0.165)	(0.113)	(0.095)
βi	4.166***	3.118***	4.155***	3.207***	3.376***
	(0.186)	(0.091)	(0.185)	(0.108)	(0.092)
var(Theta)					
βi	1	1	1	1	1
	(.)	(.)	(.)	(.)	(.)
N	4868	9165	4869	6851	11185

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: IRT analysis Inv_2010

	(1) EEDP	(2) TEPSI	(3) BDI	(4) TVIP	(5) CBCL1
Inv_Visual					
αi	1.148***	1.155***	1.156***	1.214***	1.160***
	(0.092)	(0.052)	(0.092)	(0.061)	(0.048)
βi	2.753***	1.807***	2.751***	1.810***	1.889***
	(0.087)	(0.042)	(0.088)	(0.049)	(0.039)
Inv_Talks					
αi	1.336***	1.514***	1.348***	1.551***	1.523***
	(0.092)	(0.062)	(0.092)	(0.073)	(0.058)
βi	2.392***	1.597***	2.399***	1.477***	1.746***
	(0.080)	(0.044)	(0.081)	(0.050)	(0.043)
Inv_Encourages					
αi	3.202***	3.151***	3.202***	3.068***	3.186***
	(0.461)	(0.227)	(0.462)	(0.247)	(0.219)
βi	2.520***	1.763***	2.520***	1.836***	1.919***
	(0.300)	(0.110)	(0.301)	(0.125)	(0.113)
Inv_Period					
αi	1.500***	1.699***	1.498***	1.612***	1.652***
	(0.099)	(0.071)	(0.098)	(0.076)	(0.063)
βi	-0.793***	-0.974***	-0.792***	-0.852***	-0.927***
	(0.048)	(0.038)	(0.048)	(0.041)	(0.034)
var(Theta)					
βi	1	1	1	1	1
	(.)	(.)	(.)	(.)	(.)
N	4868	9165	4869	6851	11185

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: IRT analysis Lear_Mat_2012

	(1) TADI	(2) BDI	(3) TVIP	(4) CBCL1
Mat_ColSizesShapes				
αi	2.941***	2.943***	2.962***	2.933***
	(0.088)	(0.088)	(0.090)	(0.096)
βi	0.474***	0.471***	0.510***	0.439***
	(0.042)	(0.042)	(0.042)	(0.045)
Mat_Puzzles				
αi	3.004***	3.024***	3.010***	3.030***
	(0.091)	(0.091)	(0.091)	(0.100)
βi	-0.717***	-0.721***	-0.675***	-0.721***
	(0.044)	(0.044)	(0.044)	(0.048)
Mat_Musical				
αi	1.472***	1.478***	1.461***	1.501***
	(0.044)	(0.044)	(0.044)	(0.048)
βi	1.380***	1.381***	1.399***	1.386***
	(0.034)	(0.034)	(0.034)	(0.037)
Mat_RolePlay				
αi	2.128***	2.132***	2.120***	2.109***
	(0.058)	(0.058)	(0.059)	(0.063)
βi	0.815***	0.811***	0.839***	0.809***
	(0.036)	(0.035)	(0.036)	(0.038)
Mat_Numbers				
αi	3.450***	3.433***	3.425***	3.352***
	(0.113)	(0.112)	(0.113)	(0.117)
βi	-0.814***	-0.809***	-0.768***	-0.861***
	(0.050)	(0.049)	(0.049)	(0.053)
Mat_Books				
αi	2.362***	2.361***	2.367***	2.299***
	(0.067)	(0.066)	(0.067)	(0.070)
βi	-1.442***	-1.444***	-1.420***	-1.480***
	(0.044)	(0.044)	(0.044)	(0.047)
var(Theta)				
βi	1	1	1	1
	(.)	(.)	(.)	(.)
N	10544	10665	10406	9034

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: IRT analysis Resp_2012

	(1) TADI	(2) BDI	(3) TVIP	(4) CBCL1
Resp_Vocalises				
αi	3.126***	3.083***	3.114***	3.023***
	(0.134)	(0.131)	(0.134)	(0.141)
βi	5.109***	5.058***	5.092***	5.063***
	(0.167)	(0.163)	(0.167)	(0.178)
Resp_ChildQs				
αi	3.460***	3.455***	3.481***	3.503***
	(0.166)	(0.165)	(0.168)	(0.190)
βi	5.813***	5.806***	5.844***	6.008***
	(0.217)	(0.215)	(0.221)	(0.254)
Resp_ChildSpeech				
αi	4.033***	4.050***	4.052***	3.817***
	(0.227)	(0.227)	(0.229)	(0.228)
βi	6.978***	7.009***	7.003***	6.774***
	(0.318)	(0.319)	(0.322)	(0.323)
Resp_Praises				
αi	3.169***	3.149***	3.166***	3.181***
	(0.118)	(0.115)	(0.117)	(0.131)
βi	1.906***	1.900***	1.941***	1.957***
	(0.069)	(0.068)	(0.070)	(0.078)
Resp_Caresses				
αi	2.273***	2.288***	2.301***	2.162***
	(0.069)	(0.070)	(0.071)	(0.071)
βi	0.908***	0.906***	0.914***	0.982***
	(0.040)	(0.040)	(0.041)	(0.043)
Resp_Achievement				
αi	2.586***	2.585***	2.597***	2.544***
	(0.083)	(0.082)	(0.084)	(0.089)
βi	1.523***	1.517***	1.530***	1.592***
	(0.052)	(0.051)	(0.052)	(0.057)
var(Theta)				
βi	1	1	1	1
	(.)	(.)	(.)	(.)
N	10544	10665	10406	9034

Standard errors in parentheses

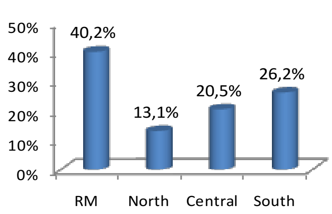
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix 4: Empirical Analysis

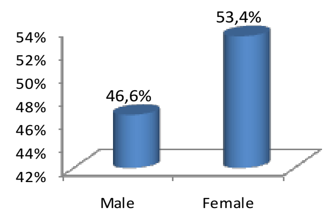
Table A4 1: Descriptive Statistics for total households for the first wave

N = 15,175 households	Mean	St. Dev.	Min	Max
People in home	4.7	1.6	2	21
Mother or caregiver`s age	29.7	7.4	14	80
Number of children < 5 years old	1.1	0.4	1	6
Income per capita	149.92	192.39	0	6,346.90

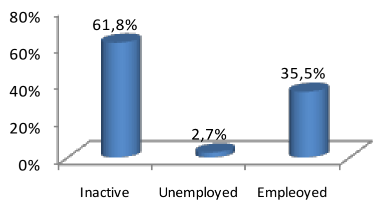
Figure A4 1: Descriptive Statistics for total households for the first wave



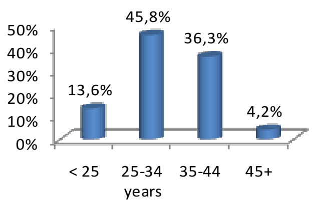
Distribution population by Geographic zone



Distribution population by Sex

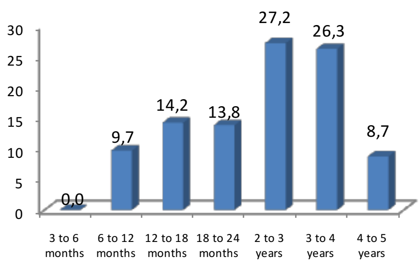


Distribution population by Occupational Status

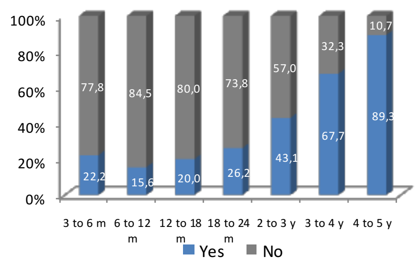


Distribution of household heads by age group

Figure A4 2: Distribution of target child for the first wave



Distribution by Age group



Distribution by Educational establishment attendance

Table A4 2: Descriptive analysis of the sample of children depending on the applied test for the first wave

		EEDP				BDI				TEPSI				TVIP				CBCL1			
		Mean	St. Dev.	Min	Max	Mean	St. Dev.	Min	Max	Mean	St. Dev.	Min	Max	Mean	St. Dev.	Min	Max	Mean	St. Dev.	Min	Max
Chil'ds test	Child's test	-0.01	0.99	-6.84	3.39	-0.01	1.00	-6.43	3.11	0.00	0.99	-5.52	4.93	-0.01	0.99	-2.14	6.73	0.00	1.00	-2.31	4.99
Demographics	Wealth q1 (poorest)	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1
	Wealth q2	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1
	Wealth q3	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1
	Wealth q4	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1
	Wealth q5 (richest)	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1
	Urban HH =1	0.90	0.30	0	1	0.90	0.30	0	1	0.90	0.30	0	1	0.90	0.30	0	1	0.89	0.31	0	1
Child's characteristics	Child’s age (mths)	16.06	4.30	7	23	16.06	4.29	7	23	38.11	8.54	24	58	41.48	6.81	30	58	35.04	10.26	18	58
	Male child =1	0.51	0.50	0	1	0.51	0.50	0	1	0.50	0.50	0	1	0.50	0.50	0	1	0.50	0.50	0	1
	Child attends ECC =1	0.19	0.39	0	1	0.19	0.39	0	1	0.57	0.50	0	1	0.64	0.48	0	1	0.51	0.50	0	1
PC's characteristics	PC’s no education =1	0.08	0.27	0	1	0.08	0.27	0	1	0.09	0.28	0	1	0.08	0.28	0	1	0.09	0.28	0	1
	PC’s Primary education =1	0.27	0.44	0	1	0.27	0.44	0	1	0.27	0.44	0	1	0.27	0.44	0	1	0.27	0.44	0	1
	PC’s Secondary education =1	0.58	0.49	0	1	0.58	0.49	0	1	0.57	0.49	0	1	0.58	0.49	0	1	0.57	0.49	0	1
	PC’s Superior education =1	0.07	0.26	0	1	0.07	0.26	0	1	0.07	0.26	0	1	0.07	0.26	0	1	0.07	0.26	0	1
	PC's age	27.92	7.10	14	69	27.93	7.11	14	69	30.22	7.36	15	71	30.52	7.39	15	71	29.85	7.35	15	71
	PC’s age <19 =1	0.16	0.36	0	1	0.16	0.36	0	1	0.15	0.36	0	1	0.15	0.36	0	1	0.15	0.36	0	1
	PC’s age 20-29 =1	0.48	0.50	0	1	0.48	0.50	0	1	0.46	0.50	0	1	0.46	0.50	0	1	0.47	0.50	0	1
	PC’s age 30-39 =1	0.31	0.46	0	1	0.31	0.46	0	1	0.33	0.47	0	1	0.33	0.47	0	1	0.32	0.47	0	1
	PC’s age 40+ =1	0.05	0.22	0	1	0.05	0.22	0	1	0.06	0.23	0	1	0.06	0.24	0	1	0.06	0.23	0	1
	PC’s LM participation =1	0.47	0.50	0	1	0.47	0.50	0	1	0.54	0.50	0	1	0.56	0.50	0	1	0.53	0.50	0	1
Family Background	Younger children =1	0.01	0.12	0	1	0.01	0.12	0	1	0.13	0.34	0	1	0.15	0.36	0	1	0.11	0.32	0	1
	Older children =1	0.53	0.50	0	1	0.53	0.50	0	1	0.55	0.50	0	1	0.54	0.50	0	1	0.55	0.50	0	1
	Two-parent family =1	0.69	0.46	0	1	0.69	0.46	0	1	0.69	0.46	0	1	0.69	0.46	0	1	0.68	0.47	0	1
Child´s Physical	Child’s length at birth+	0.10	1.01	-2.76	2.7	0.10	1.01	-2.76	2.7	0.16	1.03	-2.76	2.7	0.18	1.03	-2.76	2.7	0.15	1.02	-2.76	2.7
	Child’s weight at birth+	0.15	0.99	-3.09	2.92	0.15	0.99	-3.09	2.92	0.17	0.99	-3.09	2.96	0.19	0.98	-3.09	2.96	0.17	0.99	-3.09	2.98
PC's Cognitive	WAIS Digit Span+	50.37	9.62	6	90	50.37	9.63	6	90	50.11	9.99	6	90	50.17	9.88	6	90	50.09	9.92	6	90
	WAIS Vocabulary+	50.06	9.76	31	77	50.05	9.78	31	77	50.01	10.10	31	77	50.28	10.05	31	77	49.99	10.06	31	77
PC's Socio-emotional	BFI Extroversion	3.55	0.74	1	5	3.55	0.74	1	5	3.52	0.74	1	5	3.52	0.74	1	5	3.53	0.74	1	5
	BFI Agreeableness	3.80	0.60	1.2	5	3.80	0.60	1.2	5	3.84	0.60	1	5	3.84	0.60	1	5	3.83	0.60	1	5
	BFI Conscientiousness	3.96	0.59	1	5	3.96	0.59	1	5	4.00	0.58	1	5	4.00	0.58	1	5	3.99	0.58	1	5
	BFI Neuroticism	3.06	0.83	1	5	3.06	0.83	1	5	3.05	0.82	1	5	3.04	0.82	1	5	3.05	0.82	1	5
	BFI Openness	3.81	0.64	1.1	5	3.81	0.64	1.1	5	3.79	0.64	1	5	3.80	0.64	1	5	3.79	0.64	1	5
Home Assessment	Learning materials score+	0.00	0.80	-2.33	0.86	0.00	0.80	-2.33	0.86	0.00	0.78	-2.47	0.73	0.00	0.77	-2.51	0.71	0.00	0.78	-2.45	0.74
	Parent-Child score+	0.00	0.80	-2.96	0.82	0.00	0.80	-2.96	0.82	0.00	0.76	-2.96	0.65	0.00	0.75	-2.99	0.62	0.00	0.77	-2.96	0.67
	Responsivity score+	0.01	0.71	-2.61	0.44	0.01	0.71	-2.61	0.44	0.00	0.77	-2.56	0.59	0.00	0.77	-2.60	0.57	0.00	0.76	-2.57	0.56
	Involvement score+	0.00	0.75	-1.82	0.84	0.00	0.75	-1.82	0.84	0.00	0.79	-1.55	0.96	0.00	0.79	-1.55	0.94	0.00	0.78	-1.59	0.94

+ Standardised score

Table A4 3: Descriptive analysis of the sample of children depending on the applied test for the second wave

		TADI				BDI				TVIP				CBCL1			
		Mean	St. Dev.	Min	Max	Mean	St. Dev.	Min	Max	Mean	St. Dev.	Min	Max	Mean	St. Dev.	Min	Max
Chil'ds test	Child's test+	0.01	1.00	-10.21	4.08	0.01	0.99	-12.37	2.74	0.00	0.99	-3.50	5.85	0.01	1.00	-1.90	5.43
Demographics	Wealth q1 (poorest)	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1
	Wealth q2	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1
	Wealth q3	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1
	Wealth q4	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1
	Wealth q5 (richest)	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1	0.20	0.40	0	1
	Urban HH =1	0.89	0.31	0	1	0.89	0.31	0	1	0.89	0.31	0	1	0.89	0.31	0	1
Child's characteristics	Child’s age (mths)	56.59	12.74	33	83	56.60	12.78	33	83	56.87	12.70	33	83	53.00	10.65	33	71
	Male child =1	0.50	0.50	0	1	0.50	0.50	0	1	0.50	0.50	0	1	0.51	0.50	0	1
	Child attends ECC =1	0.86	0.35	0	1	0.86	0.35	0	1	0.87	0.34	0	1	0.83	0.37	0	1
PC's characteristics	PC’s no education =1	0.08	0.28	0	1	0.08	0.28	0	1	0.08	0.27	0	1	0.08	0.27	0	1
	PC’s Primary education =1	0.26	0.44	0	1	0.26	0.44	0	1	0.25	0.44	0	1	0.26	0.44	0	1
	PC’s Secondary education =1	0.61	0.49	0	1	0.61	0.49	0	1	0.61	0.49	0	1	0.61	0.49	0	1
	PC’s Superior education =1	0.05	0.23	0	1	0.05	0.23	0	1	0.06	0.23	0	1	0.05	0.22	0	1
	PC's age	29.55	7.43	14	71	29.56	7.43	14	71	29.58	7.42	15	71	29.19	7.35	14	70
	PC’s age <19 =1	0.15	0.36	0	1	0.15	0.36	0	1	0.15	0.36	0	1	0.15	0.36	0	1
	PC’s age 20-29 =1	0.46	0.50	0	1	0.46	0.50	0	1	0.46	0.50	0	1	0.46	0.50	0	1
	PC’s age 30-39 =1	0.33	0.47	0	1	0.33	0.47	0	1	0.33	0.47	0	1	0.33	0.47	0	1
	PC’s age 40+ =1	0.06	0.24	0	1	0.06	0.24	0	1	0.06	0.24	0	1	0.06	0.23	0	1
	PC’s LM participation =1	0.54	0.50	0	1	0.54	0.50	0	1	0.54	0.50	0	1	0.54	0.50	0	1
Family Background	Younger children =1	0.21	0.41	0	1	0.21	0.41	0	1	0.21	0.41	0	1	0.20	0.40	0	1
	Older children =1	0.55	0.50	0	1	0.55	0.50	0	1	0.55	0.50	0	1	0.55	0.50	0	1
	Two-parent family =1	0.67	0.47	0	1	0.67	0.47	0	1	0.67	0.47	0	1	0.67	0.47	0	1
Child´s Physical	Child’s length at birth+	0.14	1.03	-2.76	2.7	0.14	1.02	-2.76	2.7	0.14	1.03	-2.76	2.7	0.13	1.02	-2.76	2.7
	Child’s weight at birth+	0.17	0.99	-3.09	2.98	0.17	1.00	-3.09	2.98	0.17	0.99	-3.09	2.98	0.16	1.00	-3.09	2.98
PC's Cognitive	WAIS Digit Span+	49.91	9.82	6	90	49.92	9.82	6	90	49.95	9.82	6	90	49.96	9.83	6	90
	WAIS Vocabulary+	49.75	9.92	31	77	49.75	9.94	31	77	49.79	9.92	31	77	49.70	9.92	31	77
PC's Socio-emotional	BFI Extroversion	3.52	0.74	1	5	3.52	0.74	1	5	3.53	0.74	1	5	3.52	0.74	1	5
	BFI Agreeableness	3.83	0.60	1	5	3.83	0.60	1	5	3.83	0.60	1	5	3.82	0.60	1	5
	BFI Conscientiousness	3.99	0.58	1	5	3.98	0.58	1	5	3.99	0.58	1	5	3.98	0.58	1	5
	BFI Neuroticism	3.06	0.82	1	5	3.06	0.82	1	5	3.06	0.82	1	5	3.07	0.82	1	5
	BFI Openness	3.79	0.64	1	5	3.79	0.64	1	5	3.80	0.64	1	5	3.79	0.64	1	5
Home Assessment	Learning materials score+	0.00	0.88	-1.34	1.31	0.00	0.88	-1.34	1.31	0.00	0.88	-1.35	1.30	0.00	0.88	-1.34	1.32
	Parent-Child score+	0.01	0.78	-2.54	0.90	0.01	0.78	-2.54	0.90	0.01	0.77	-2.54	0.89	0.01	0.77	-2.58	0.89
	Responsivity score+	0.00	0.82	-2.24	0.72	0.00	0.82	-2.24	0.72	0.00	0.82	-2.24	0.72	0.00	0.81	-2.27	0.71

+ Standardised score

Table A.8: Children's cognitive eedp test z-score: Linear regression model

	(1)	(2)	(3)	(4)	(5)
	eedp+	eedp+	eedp+	eedp+	eedp+
Wealth q2	0.124***	0.066	0.066	0.033	0.011
	(0.047)	(0.052)	(0.052)	(0.052)	(0.050)
Wealth q3	0.147***	0.107**	0.110**	0.062	0.029
	(0.047)	(0.054)	(0.054)	(0.053)	(0.053)
Wealth q4	0.210***	0.165***	0.168***	0.088	0.040
	(0.046)	(0.054)	(0.054)	(0.054)	(0.053)
Wealth q5	0.276***	0.191***	0.197***	0.091	0.035
	(0.046)	(0.056)	(0.055)	(0.056)	(0.056)
Child's age (mths)	0.016***	0.017***	0.017***	0.017***	0.015***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)
Male child =1	-0.148***	-0.158***	-0.156***	-0.156***	-0.145***
	(0.028)	(0.031)	(0.031)	(0.031)	(0.031)
Child attends ECC =1		-0.090**	-0.090**	-0.093**	-0.086**
		(0.043)	(0.043)	(0.043)	(0.042)
Urban HH =1		0.090*	0.093*	0.099*	0.089*
		(0.052)	(0.052)	(0.051)	(0.051)
PC's Primary education =1		0.070	0.074	-0.016	-0.052
		(0.071)	(0.071)	(0.071)	(0.071)
PC's Secondary education =1		0.132*	0.135*	-0.010	-0.065
		(0.070)	(0.070)	(0.072)	(0.072)
PC's Superior education =1		0.108	0.109	-0.116	-0.155
		(0.093)	(0.092)	(0.096)	(0.096)
PC's age 20-29 =1		-0.079	-0.090*	-0.137***	-0.146***
		(0.049)	(0.049)	(0.049)	(0.048)
PC's age 30-39 =1		-0.134**	-0.147**	-0.214***	-0.235***
		(0.058)	(0.058)	(0.059)	(0.058)
PC's age 40+ =1		-0.056	-0.067	-0.120	-0.155*
		(0.088)	(0.087)	(0.087)	(0.086)
Younger children =1		-0.015	-0.000	-0.016	0.016
		(0.119)	(0.117)	(0.112)	(0.112)
Older children =1		-0.059	-0.066*	-0.054	-0.040
		(0.037)	(0.037)	(0.036)	(0.036)
Two-parent family =1		-0.020	-0.015	-0.020	-0.021
		(0.036)	(0.036)	(0.036)	(0.035)
PC's LM participation =1		0.023	0.023	0.012	0.008
		(0.034)	(0.034)	(0.034)	(0.034)
Child's length at birth+			0.023	0.020	0.025
			(0.020)	(0.020)	(0.020)
Child's weight at birth+			0.051**	0.050**	0.046**
			(0.021)	(0.021)	(0.021)
WAIS_Digit Span+				0.005***	0.006***
				(0.002)	(0.002)
WAIS_Vocabulary+				0.006***	0.002
				(0.002)	(0.002)
BFI_Extroversion				0.046**	0.026
				(0.023)	(0.023)
BFI_Agreeableness				0.015	0.011
				(0.029)	(0.029)
BFI_Conscientiousness				0.059*	0.036
				(0.030)	(0.031)
BFI_Neuroticism				-0.007	-0.008
				(0.022)	(0.021)
BFI_Openness				0.103***	0.077***
				(0.028)	(0.028)
Learning materials score+					0.035
					(0.023)
Parent-Child score+					0.044**
					(0.021)
Responsivity score+					0.132***
					(0.025)
Involvement score+					0.114***
					(0.024)
Constant	-0.339***	-0.378***	-0.381***	-1.543***	-0.962***
	(0.063)	(0.104)	(0.104)	(0.232)	(0.243)
N	4868	4065	4065	3964	3962
R^2	0.0190	0.0258	0.0304	0.0513	0.0739

Standard errors in parentheses
"Robust standard errors in parentheses; + z-score"
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Children's cognitive bdi test z-score: Linear regression model

	(1)	(2)	(3)	(4)	(5)
	bdi+	bdi+	bdi+	bdi+	bdi+
Wealth q2	0.150***	0.087	0.086	0.040	0.009
	(0.047)	(0.053)	(0.053)	(0.052)	(0.051)
Wealth q3	0.231***	0.161***	0.163***	0.092*	0.050
	(0.047)	(0.055)	(0.055)	(0.054)	(0.054)
Wealth q4	0.352***	0.243***	0.245***	0.135**	0.077
	(0.046)	(0.056)	(0.056)	(0.055)	(0.054)
Wealth q5	0.390***	0.265***	0.270***	0.127**	0.057
	(0.046)	(0.057)	(0.057)	(0.057)	(0.056)
Child's age (mths)	0.013***	0.013***	0.013***	0.013***	0.010***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Male child =1	-0.166***	-0.185***	-0.182***	-0.180***	-0.165***
	(0.028)	(0.031)	(0.031)	(0.031)	(0.030)
Child attends ECC =1		-0.094**	-0.095**	-0.098**	-0.089**
		(0.043)	(0.042)	(0.042)	(0.041)
Urban HH =1		0.118**	0.122**	0.131**	0.117**
		(0.057)	(0.057)	(0.055)	(0.054)
PC's Primary education =1		0.196***	0.200***	0.100	0.052
		(0.071)	(0.071)	(0.070)	(0.069)
PC's Secondary education =1		0.256***	0.259***	0.073	0.002
		(0.069)	(0.069)	(0.071)	(0.070)
PC's Superior education =1		0.254***	0.254***	-0.054	-0.109
		(0.092)	(0.091)	(0.094)	(0.093)
PC's age 20-29 =1		-0.030	-0.042	-0.075	-0.088*
		(0.050)	(0.050)	(0.050)	(0.049)
PC's age 30-39 =1		-0.089	-0.102*	-0.171***	-0.196***
		(0.059)	(0.059)	(0.060)	(0.059)
PC's age 40+ =1		-0.120	-0.132	-0.191**	-0.236**
		(0.094)	(0.094)	(0.093)	(0.092)
Younger children =1		0.142	0.157	0.142	0.176
		(0.123)	(0.123)	(0.117)	(0.111)
Older children =1		-0.070*	-0.078**	-0.072**	-0.055
		(0.037)	(0.037)	(0.036)	(0.035)
Two-parent family =1		0.037	0.042	0.029	0.026
		(0.036)	(0.036)	(0.036)	(0.035)
PC's LM participation =1		0.053	0.053	0.054	0.050
		(0.034)	(0.034)	(0.034)	(0.034)
Child's length at birth+			0.030	0.029	0.036*
			(0.020)	(0.020)	(0.019)
Child's weight at birth+			0.049**	0.048**	0.042**
			(0.020)	(0.020)	(0.020)
WAIS_Digit Span+				-0.001	-0.001
				(0.002)	(0.002)
WAIS_Vocabulary+				0.014***	0.010***
				(0.002)	(0.002)
BFI_Extroversion				0.044*	0.016
				(0.023)	(0.022)
BFI_Agreeableness				0.012	0.006
				(0.028)	(0.028)
BFI_Conscientiousness				0.058*	0.028
				(0.031)	(0.030)
BFI_Neuroticism				0.005	0.005
				(0.021)	(0.021)
BFI_Openness				0.156***	0.123***
				(0.027)	(0.027)
Learning materials score+					0.039*
					(0.022)
Parent-Child score+					0.090***
					(0.020)
Responsivity score+					0.196***
					(0.025)
Involvement score+					0.095***
					(0.023)
Constant	-0.358***	-0.553***	-0.556***	-1.984***	-1.230***
	(0.065)	(0.108)	(0.107)	(0.232)	(0.239)
N	4869	4065	4065	3963	3961
R^2	0.0304	0.0402	0.0454	0.0822	0.1180

Standard errors in parentheses
"Robust standard errors in parentheses; + z-score"
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.10: Children's cognitive tepsi test z-score: Linear regression model

	(1)	(2)	(3)	(4)	(5)
	tepsi+	tepsi+	tepsi+	tepsi+	tepsi+
Wealth q2	0.213***	0.167***	0.167***	0.146***	0.109***
	(0.033)	(0.037)	(0.037)	(0.037)	(0.036)
Wealth q3	0.357***	0.282***	0.280***	0.220***	0.158***
	(0.033)	(0.037)	(0.037)	(0.037)	(0.037)
Wealth q4	0.432***	0.298***	0.300***	0.224***	0.156***
	(0.033)	(0.039)	(0.039)	(0.040)	(0.039)
Wealth q5	0.649***	0.455***	0.460***	0.326***	0.239***
	(0.033)	(0.042)	(0.042)	(0.043)	(0.043)
Child's age (mths)	0.005***	0.001	0.001	0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male child =1	-0.287***	-0.294***	-0.291***	-0.289***	-0.281***
	(0.020)	(0.022)	(0.022)	(0.022)	(0.022)
Child attends ECC =1		0.188***	0.188***	0.193***	0.192***
		(0.026)	(0.026)	(0.026)	(0.025)
Urban HH =1		0.127***	0.127***	0.120***	0.090**
		(0.038)	(0.038)	(0.037)	(0.036)
PC's Primary education =1		0.190***	0.192***	0.117**	0.076*
		(0.046)	(0.046)	(0.047)	(0.046)
PC's Secondary education =1		0.337***	0.343***	0.189***	0.134***
		(0.046)	(0.046)	(0.048)	(0.046)
PC's Superior education =1		0.510***	0.520***	0.254***	0.193***
		(0.064)	(0.064)	(0.068)	(0.066)
PC's age 20-29 =1		0.011	0.006	-0.003	-0.016
		(0.035)	(0.035)	(0.035)	(0.034)
PC's age 30-39 =1		0.019	0.013	-0.005	-0.025
		(0.041)	(0.041)	(0.041)	(0.040)
PC's age 40+ =1		-0.017	-0.018	-0.018	-0.043
		(0.060)	(0.060)	(0.060)	(0.059)
Younger children =1		-0.076**	-0.077**	-0.104***	-0.061*
		(0.033)	(0.032)	(0.033)	(0.032)
Older children =1		-0.123***	-0.129***	-0.130***	-0.094***
		(0.027)	(0.027)	(0.027)	(0.026)
Two-parent family =1		0.058**	0.054**	0.049*	0.034
		(0.026)	(0.026)	(0.026)	(0.026)
PC's LM participation =1		0.110***	0.112***	0.086***	0.083***
		(0.024)	(0.024)	(0.024)	(0.023)
Child's length at birth+			0.034**	0.033**	0.036***
			(0.014)	(0.014)	(0.013)
Child's weight at birth+			0.023	0.021	0.016
			(0.014)	(0.014)	(0.014)
WAIS_Digit Span+				0.010***	0.010***
				(0.001)	(0.001)
WAIS_Vocabulary+				0.009***	0.005***
				(0.001)	(0.001)
BFI_Extroversion				0.006	-0.022
				(0.016)	(0.016)
BFI_Agreeableness				-0.007	-0.016
				(0.021)	(0.020)
BFI_Conscientiousness				0.012	-0.014
				(0.022)	(0.022)
BFI_Neuroticism				-0.038**	-0.030**
				(0.016)	(0.015)
BFI_Openness				0.003	-0.031
				(0.019)	(0.019)
Learning materials score+					0.047***
					(0.017)
Parent-Child score+					0.055***
					(0.016)
Responsivity score+					0.247***
					(0.016)
Involvement score+					0.097***
					(0.016)
Constant	-0.367***	-0.661***	-0.658***	-1.339***	-0.618***
	(0.049)	(0.076)	(0.076)	(0.172)	(0.173)
N	9167	7335	7335	7167	7158
R^2	0.0722	0.1178	0.1205	0.1427	0.1949

Standard errors in parentheses
"Robust standard errors in parentheses; + z-score"
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.11: Children's cognitive tvip test z-score: Linear regression model

	(1)	(2)	(3)	(4)	(5)
	tvip+	tvip+	tvip+	tvip+	tvip+
Wealth q2	0.207***	0.176***	0.178***	0.139***	0.110***
	(0.035)	(0.038)	(0.038)	(0.038)	(0.038)
Wealth q3	0.395***	0.294***	0.296***	0.226***	0.175***
	(0.036)	(0.040)	(0.040)	(0.040)	(0.041)
Wealth q4	0.509***	0.358***	0.361***	0.264***	0.213***
	(0.036)	(0.042)	(0.042)	(0.043)	(0.044)
Wealth q5	0.912***	0.629***	0.633***	0.462***	0.395***
	(0.042)	(0.048)	(0.048)	(0.048)	(0.049)
Child's age (mths)	0.014***	0.014***	0.014***	0.013***	0.013***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Male child =1	-0.067***	-0.041	-0.040	-0.039	-0.038
	(0.025)	(0.027)	(0.027)	(0.027)	(0.027)
Child attends ECC =1		0.025	0.025	0.028	0.033
		(0.031)	(0.031)	(0.031)	(0.031)
Urban HH =1		0.123***	0.121***	0.123***	0.109***
		(0.041)	(0.041)	(0.040)	(0.040)
PC's Primary education =1		0.115**	0.116**	0.037	0.009
		(0.046)	(0.046)	(0.048)	(0.048)
PC's Secondary education =1		0.330***	0.333***	0.152***	0.113**
		(0.047)	(0.047)	(0.050)	(0.050)
PC's Superior education =1		0.644***	0.651***	0.340***	0.285***
		(0.087)	(0.087)	(0.089)	(0.091)
PC's age 20-29 =1		0.068*	0.064	0.053	0.050
		(0.040)	(0.040)	(0.040)	(0.040)
PC's age 30-39 =1		0.097**	0.094**	0.068	0.062
		(0.048)	(0.048)	(0.048)	(0.048)
PC's age 40+ =1		0.210***	0.211***	0.193**	0.181**
		(0.078)	(0.078)	(0.079)	(0.079)
Younger children =1		-0.131***	-0.131***	-0.155***	-0.135***
		(0.038)	(0.038)	(0.038)	(0.038)
Older children =1		-0.154***	-0.159***	-0.157***	-0.130***
		(0.033)	(0.033)	(0.032)	(0.032)
Two-parent family =1		0.105***	0.103***	0.108***	0.098***
		(0.030)	(0.030)	(0.030)	(0.030)
PC's LM participation =1		0.049*	0.050*	0.019	0.021
		(0.028)	(0.028)	(0.028)	(0.028)
Child's length at birth+			-0.004	-0.007	-0.005
			(0.018)	(0.018)	(0.018)
Child's weight at birth+			0.035*	0.032*	0.030
			(0.019)	(0.019)	(0.019)
WAIS_Digit Span+				0.011***	0.011***
				(0.002)	(0.002)
WAIS_Vocabulary+				0.012***	0.010***
				(0.002)	(0.002)
BFI_Extroversion				0.052***	0.038**
				(0.019)	(0.019)
BFI_Agreeableness				0.003	-0.005
				(0.023)	(0.023)
BFI_Conscientiousness				0.009	-0.008
				(0.025)	(0.025)
BFI_Neuroticism				-0.030*	-0.025
				(0.018)	(0.018)
BFI_Openness				-0.018	-0.039
				(0.024)	(0.024)
Learning materials score+					0.070***
					(0.018)
Parent-Child score+					0.068***
					(0.019)
Responsivity score+					0.085***
					(0.019)
Involvement score+					0.046**
					(0.021)
Constant	-0.934***	-1.316***	-1.314***	-2.293***	-1.905***
	(0.082)	(0.108)	(0.107)	(0.199)	(0.207)
N	6852	5510	5510	5387	5381
R^2	0.1070	0.1425	0.1435	0.1746	0.1875

Standard errors in parentheses
"Robust standard errors in parentheses; + z-score"
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.12: Children's cognitive cbcl1 test z-score: Linear regression model

	(1)	(2)	(3)	(4)	(5)
	cbcl1+	cbcl1+	cbcl1+	cbcl1+	cbcl1+
Wealth q2	-0.078**	-0.009	-0.011	-0.009	0.008
	(0.032)	(0.036)	(0.036)	(0.033)	(0.033)
Wealth q3	-0.266***	-0.166***	-0.167***	-0.114***	-0.079**
	(0.031)	(0.036)	(0.036)	(0.034)	(0.034)
Wealth q4	-0.359***	-0.209***	-0.211***	-0.130***	-0.096***
	(0.030)	(0.037)	(0.037)	(0.035)	(0.035)
Wealth q5	-0.687***	-0.430***	-0.433***	-0.257***	-0.210***
	(0.030)	(0.038)	(0.038)	(0.036)	(0.037)
Child's age (mths)	-0.000	-0.002	-0.002	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male child =1	0.109***	0.101***	0.102***	0.095***	0.091***
	(0.019)	(0.021)	(0.021)	(0.019)	(0.019)
Child attends ECC =1		0.100***	0.100***	0.064***	0.064***
		(0.023)	(0.023)	(0.022)	(0.022)
Urban HH =1		0.139***	0.140***	0.096***	0.109***
		(0.036)	(0.036)	(0.034)	(0.034)
PC's Primary education =1		-0.127***	-0.129***	-0.058	-0.035
		(0.048)	(0.048)	(0.046)	(0.046)
PC's Secondary education =1		-0.369***	-0.371***	-0.170***	-0.140***
		(0.047)	(0.047)	(0.046)	(0.046)
PC's Superior education =1		-0.591***	-0.594***	-0.256***	-0.224***
		(0.060)	(0.060)	(0.058)	(0.058)
PC's age 20-29 =1		-0.168***	-0.166***	-0.124***	-0.119***
		(0.033)	(0.033)	(0.031)	(0.031)
PC's age 30-39 =1		-0.309***	-0.307***	-0.194***	-0.186***
		(0.038)	(0.038)	(0.037)	(0.037)
PC's age 40+ =1		-0.412***	-0.413***	-0.277***	-0.262***
		(0.054)	(0.054)	(0.051)	(0.051)
Younger children =1		0.145***	0.145***	0.159***	0.143***
		(0.034)	(0.034)	(0.031)	(0.031)
Older children =1		0.042*	0.045*	0.029	0.017
		(0.025)	(0.025)	(0.023)	(0.023)
Two-parent family =1		-0.096***	-0.094***	-0.081***	-0.073***
		(0.024)	(0.024)	(0.022)	(0.022)
PC's LM participation =1		-0.053**	-0.053**	-0.007	-0.006
		(0.022)	(0.022)	(0.021)	(0.021)
Child's length at birth+			0.009	0.010	0.009
			(0.013)	(0.012)	(0.012)
Child's weight at birth+			-0.026*	-0.027**	-0.023*
			(0.014)	(0.013)	(0.013)
WAIS_Digit Span+				-0.007***	-0.007***
				(0.001)	(0.001)
WAIS_Vocabulary+				-0.004***	-0.002
				(0.001)	(0.001)
BFI_Extroversion				-0.040***	-0.026*
				(0.014)	(0.014)
BFI_Agreeableness				-0.097***	-0.094***
				(0.018)	(0.018)
BFI_Conscientiousness				-0.120***	-0.108***
				(0.019)	(0.019)
BFI_Neuroticism				0.372***	0.367***
				(0.013)	(0.013)
BFI_Openness				0.039**	0.053***
				(0.017)	(0.017)
Learning materials score+					-0.033**
					(0.015)
Parent-Child score+					-0.036**
					(0.014)
Responsivity score+					-0.130***
					(0.015)
Involvement score+					-0.001
					(0.014)
Constant	0.221***	0.539***	0.539***	0.480***	0.182
	(0.040)	(0.071)	(0.071)	(0.148)	(0.151)
N	11193	8867	8867	8664	8650
R^2	0.0635	0.1043	0.1047	0.2346	0.2452

Standard errors in parentheses
"Robust standard errors in parentheses; + z-score"
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.13: Children's cognitive tadi test z-score: Linear regression model

	(1)	(2)	(3)	(4)	(5)
	tadi+	tadi+	tadi+	tadi+	tadi+
Wealth q2	0.263***	0.174***	0.174***	0.154***	0.124***
	(0.031)	(0.030)	(0.030)	(0.030)	(0.030)
Wealth q3	0.346***	0.200***	0.200***	0.158***	0.109***
	(0.031)	(0.032)	(0.032)	(0.032)	(0.032)
Wealth q4	0.492***	0.298***	0.300***	0.247***	0.179***
	(0.030)	(0.032)	(0.032)	(0.032)	(0.032)
Wealth q5	0.703***	0.450***	0.454***	0.368***	0.280***
	(0.031)	(0.033)	(0.033)	(0.034)	(0.034)
Child's age (mths)	0.001	-0.002**	-0.002**	-0.002**	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male child =1	-0.185***	-0.190***	-0.188***	-0.188***	-0.168***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Child attends ECC =1		0.286***	0.287***	0.274***	0.267***
		(0.034)	(0.034)	(0.033)	(0.033)
Urban HH =1		0.114***	0.115***	0.108***	0.077**
		(0.031)	(0.031)	(0.031)	(0.031)
PC's Primary education =1		0.256***	0.258***	0.179***	0.156***
		(0.042)	(0.042)	(0.042)	(0.042)
PC's Secondary education =1		0.448***	0.450***	0.321***	0.264***
		(0.041)	(0.041)	(0.041)	(0.041)
PC's Superior education =1		0.548***	0.550***	0.350***	0.287***
		(0.061)	(0.061)	(0.062)	(0.063)
PC's age 20-29 =1		0.037	0.035	0.018	0.006
		(0.028)	(0.028)	(0.029)	(0.029)
PC's age 30-39 =1		0.064*	0.062*	0.027	0.010
		(0.033)	(0.033)	(0.034)	(0.034)
PC's age 40+ =1		0.033	0.033	-0.004	-0.024
		(0.050)	(0.050)	(0.051)	(0.051)
Younger children =1		-0.006	-0.007	-0.012	0.017
		(0.023)	(0.023)	(0.023)	(0.023)
Older children =1		-0.110***	-0.116***	-0.119***	-0.111***
		(0.023)	(0.023)	(0.022)	(0.023)
Two-parent family =1		0.041**	0.040*	0.032	0.030
		(0.021)	(0.021)	(0.021)	(0.021)
PC's LM participation =1		0.073***	0.073***	0.061***	0.060***
		(0.020)	(0.020)	(0.020)	(0.019)
Child's length at birth+			0.022*	0.020*	0.020*
			(0.011)	(0.012)	(0.011)
Child's weight at birth+			0.020	0.018	0.016
			(0.012)	(0.012)	(0.012)
WAIS_Digit Span+				0.005***	0.005***
				(0.001)	(0.001)
WAIS_Vocabulary+				0.008***	0.006***
				(0.001)	(0.001)
BFI_Extroversion				0.021	0.009
				(0.014)	(0.014)
BFI_Agreeableness				0.017	0.017
				(0.017)	(0.017)
BFI_Conscientiousness				0.051***	0.035*
				(0.019)	(0.019)
BFI_Neuroticism				0.006	0.013
				(0.012)	(0.012)
BFI_Openness				0.011	-0.003
				(0.017)	(0.017)
Learning materials score+					0.146***
					(0.012)
Parent-Child score+					0.028**
					(0.012)
Responsivity score+					0.106***
					(0.012)
Constant	-0.309***	-0.776***	-0.775***	-1.621***	-1.255***
	(0.048)	(0.065)	(0.065)	(0.136)	(0.136)
N	10865	10865	10865	10577	10228
R^2	0.0649	0.1006	0.1020	0.1142	0.1434

Standard errors in parentheses
"Robust standard errors in parentheses; + z-score"
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.14: Children's cognitive bdi test z-score: Linear regression model

	(1)	(2)	(3)	(4)	(5)
	bdi+	bdi+	bdi+	bdi+	bdi+
Wealth q2	0.209***	0.134***	0.134***	0.120***	0.089***
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
Wealth q3	0.303***	0.177***	0.178***	0.140***	0.090***
	(0.031)	(0.032)	(0.032)	(0.033)	(0.033)
Wealth q4	0.372***	0.202***	0.204***	0.160***	0.092***
	(0.030)	(0.032)	(0.032)	(0.032)	(0.033)
Wealth q5	0.581***	0.344***	0.347***	0.278***	0.187***
	(0.030)	(0.034)	(0.033)	(0.034)	(0.035)
Child's age (mths)	0.001	-0.001	-0.001	-0.001*	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male child =1	-0.293***	-0.295***	-0.293***	-0.298***	-0.283***
	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)
Child attends ECC =1		0.203***	0.203***	0.196***	0.191***
		(0.032)	(0.032)	(0.031)	(0.032)
Urban HH =1		0.089***	0.090***	0.089***	0.054*
		(0.031)	(0.031)	(0.031)	(0.031)
PC's Primary education =1		0.252***	0.253***	0.188***	0.163***
		(0.042)	(0.042)	(0.042)	(0.042)
PC's Secondary education =1		0.410***	0.412***	0.292***	0.234***
		(0.041)	(0.041)	(0.042)	(0.041)
PC's Superior education =1		0.565***	0.567***	0.377***	0.320***
		(0.056)	(0.056)	(0.057)	(0.057)
PC's age 20-29 =1		0.016	0.015	-0.011	-0.015
		(0.028)	(0.028)	(0.029)	(0.029)
PC's age 30-39 =1		0.060*	0.059*	0.011	-0.006
		(0.033)	(0.033)	(0.034)	(0.034)
PC's age 40+ =1		0.050	0.050	0.002	-0.009
		(0.047)	(0.047)	(0.048)	(0.048)
Younger children =1		0.004	0.004	0.008	0.030
		(0.023)	(0.023)	(0.023)	(0.023)
Older children =1		-0.072***	-0.076***	-0.078***	-0.063***
		(0.022)	(0.022)	(0.022)	(0.022)
Two-parent family =1		0.066***	0.065***	0.057***	0.048**
		(0.022)	(0.022)	(0.021)	(0.021)
PC's LM participation =1		0.066***	0.065***	0.052***	0.050**
		(0.020)	(0.020)	(0.020)	(0.019)
Child's length at birth+			0.015	0.014	0.013
			(0.012)	(0.012)	(0.012)
Child's weight at birth+			0.013	0.011	0.012
			(0.012)	(0.012)	(0.012)
WAIS_Digit Span+				0.005***	0.005***
				(0.001)	(0.001)
WAIS_Vocabulary+				0.004***	0.002*
				(0.001)	(0.001)
BFI_Extroversion				0.034**	0.025*
				(0.013)	(0.013)
BFI_Agreeableness				0.024	0.022
				(0.018)	(0.017)
BFI_Conscientiousness				0.072***	0.059***
				(0.018)	(0.018)
BFI_Neuroticism				-0.023*	-0.018
				(0.013)	(0.013)
BFI_Openness				0.044***	0.030*
				(0.017)	(0.016)
Learning materials score+					0.145***
					(0.012)
Parent-Child score+					0.022*
					(0.012)
Responsivity score+					0.109***
					(0.012)
Constant	-0.176***	-0.611***	-0.610***	-1.473***	-1.122***
	(0.047)	(0.066)	(0.066)	(0.140)	(0.137)
N	10909	10909	10909	10616	10343
R^2	0.0593	0.0850	0.0856	0.0989	0.1289

Standard errors in parentheses
"Robust standard errors in parentheses; + z-score"
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.15: Children's cognitive tvip test z-score: Linear regression model

	(1)	(2)	(3)	(4)	(5)
	tvip+	tvip+	tvip+	tvip+	tvip+
Wealth q2	0.217***	0.132***	0.132***	0.116***	0.102***
	(0.029)	(0.029)	(0.029)	(0.029)	(0.030)
Wealth q3	0.380***	0.231***	0.231***	0.191***	0.168***
	(0.030)	(0.031)	(0.031)	(0.032)	(0.032)
Wealth q4	0.507***	0.304***	0.305***	0.250***	0.214***
	(0.031)	(0.033)	(0.033)	(0.034)	(0.034)
Wealth q5	0.778***	0.477***	0.481***	0.387***	0.342***
	(0.032)	(0.035)	(0.035)	(0.036)	(0.037)
Child's age (mths)	0.002***	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male child =1	-0.124***	-0.122***	-0.121***	-0.122***	-0.124***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Child attends ECC =1		0.135***	0.135***	0.129***	0.120***
		(0.031)	(0.031)	(0.031)	(0.032)
Urban HH =1		0.138***	0.138***	0.125***	0.110***
		(0.031)	(0.031)	(0.030)	(0.031)
PC's Primary education =1		0.225***	0.227***	0.168***	0.153***
		(0.035)	(0.035)	(0.036)	(0.036)
PC's Secondary education =1		0.434***	0.436***	0.322***	0.295***
		(0.035)	(0.035)	(0.037)	(0.037)
PC's Superior education =1		0.643***	0.645***	0.456***	0.420***
		(0.060)	(0.060)	(0.064)	(0.065)
PC's age 20-29 =1		0.093***	0.092***	0.071**	0.068**
		(0.029)	(0.029)	(0.029)	(0.030)
PC's age 30-39 =1		0.231***	0.229***	0.189***	0.183***
		(0.034)	(0.034)	(0.034)	(0.035)
PC's age 40+ =1		0.257***	0.257***	0.231***	0.226***
		(0.052)	(0.052)	(0.052)	(0.053)
Younger children =1		-0.073***	-0.074***	-0.073***	-0.066***
		(0.024)	(0.024)	(0.024)	(0.024)
Older children =1		-0.184***	-0.188***	-0.188***	-0.183***
		(0.022)	(0.022)	(0.022)	(0.023)
Two-parent family =1		0.094***	0.093***	0.088***	0.091***
		(0.021)	(0.021)	(0.022)	(0.022)
PC's LM participation =1		0.041**	0.040**	0.025	0.025
		(0.020)	(0.020)	(0.020)	(0.021)
Child's length at birth+			0.019	0.018	0.014
			(0.012)	(0.012)	(0.012)
Child's weight at birth+			0.014	0.015	0.017
			(0.012)	(0.012)	(0.012)
WAIS_Digit Span+				0.003**	0.003***
				(0.001)	(0.001)
WAIS_Vocabulary+				0.011***	0.010***
				(0.001)	(0.001)
BFI_Extroversion				0.033**	0.027*
				(0.014)	(0.014)
BFI_Agreeableness				-0.010	-0.004
				(0.018)	(0.018)
BFI_Conscientiousness				0.014	0.009
				(0.018)	(0.019)
BFI_Neuroticism				-0.021	-0.015
				(0.013)	(0.013)
BFI_Openness				-0.000	-0.003
				(0.017)	(0.017)
Learning materials score+					0.090***
					(0.012)
Parent-Child score+					0.025*
					(0.013)
Responsivity score+					0.023*
					(0.012)
Constant	-0.423***	-0.929***	-0.929***	-1.490***	-1.360***
	(0.048)	(0.064)	(0.064)	(0.143)	(0.147)
N	10708	10708	10708	10419	10092
R^2	0.0747	0.1117	0.1126	0.1264	0.1359

Standard errors in parentheses
"Robust standard errors in parentheses; + z-score"
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.16: Children's cognitive cbcl1 test z-score: Linear regression model

	(1)	(2)	(3)	(4)	(5)
	cbcl1+	cbcl1+	cbcl1+	cbcl1+	cbcl1+
Wealth q2	-0.121***	-0.070**	-0.069**	-0.045	-0.027
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)
Wealth q3	-0.169***	-0.072**	-0.071**	-0.023	0.013
	(0.035)	(0.036)	(0.036)	(0.036)	(0.036)
Wealth q4	-0.280***	-0.150***	-0.149***	-0.079**	-0.036
	(0.034)	(0.036)	(0.036)	(0.036)	(0.036)
Wealth q5	-0.522***	-0.308***	-0.307***	-0.189***	-0.134***
	(0.033)	(0.037)	(0.037)	(0.037)	(0.038)
Child's age (mths)	-0.002*	-0.004***	-0.004***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male child =1	0.115***	0.110***	0.110***	0.106***	0.095***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Child attends ECC =1		0.123***	0.123***	0.121***	0.139***
		(0.030)	(0.030)	(0.030)	(0.031)
Urban HH =1		0.056	0.056	0.034	0.058*
		(0.035)	(0.035)	(0.034)	(0.034)
PC's Primary education =1		-0.158***	-0.157***	-0.097*	-0.083*
		(0.049)	(0.049)	(0.050)	(0.050)
PC's Secondary education =1		-0.366***	-0.366***	-0.223***	-0.187***
		(0.048)	(0.048)	(0.050)	(0.050)
PC's Superior education =1		-0.576***	-0.575***	-0.323***	-0.289***
		(0.062)	(0.062)	(0.064)	(0.065)
PC's age 20-29 =1		-0.056*	-0.057*	-0.010	-0.001
		(0.033)	(0.033)	(0.033)	(0.033)
PC's age 30-39 =1		-0.134***	-0.134***	-0.031	-0.004
		(0.038)	(0.038)	(0.038)	(0.039)
PC's age 40+ =1		-0.121**	-0.120**	-0.010	-0.003
		(0.054)	(0.054)	(0.054)	(0.054)
Younger children =1		0.074***	0.075***	0.075***	0.057**
		(0.027)	(0.027)	(0.027)	(0.027)
Older children =1		-0.009	-0.011	-0.018	-0.025
		(0.024)	(0.024)	(0.023)	(0.024)
Two-parent family =1		-0.136***	-0.137***	-0.117***	-0.113***
		(0.023)	(0.023)	(0.023)	(0.024)
PC's LM participation =1		-0.042*	-0.041*	-0.023	-0.017
		(0.022)	(0.022)	(0.021)	(0.021)
Child's length at birth+			-0.021*	-0.018	-0.014
			(0.013)	(0.013)	(0.013)
Child's weight at birth+			0.022*	0.023*	0.022*
			(0.013)	(0.013)	(0.013)
WAIS_Digit Span+				-0.002**	-0.002**
				(0.001)	(0.001)
WAIS_Vocabulary+				-0.007***	-0.006***
				(0.001)	(0.001)
BFI_Extroversion				-0.017	-0.009
				(0.015)	(0.015)
BFI_Agreeableness				-0.087***	-0.084***
				(0.019)	(0.019)
BFI_Conscientiousness				-0.070***	-0.067***
				(0.019)	(0.019)
BFI_Neuroticism				0.204***	0.200***
				(0.014)	(0.014)
BFI_Openness				0.011	0.020
				(0.018)	(0.018)
Learning materials score+					-0.076***
					(0.013)
Parent-Child score+					-0.016
					(0.014)
Responsivity score+					-0.099***
					(0.014)
Constant	0.265***	0.613***	0.611***	0.844***	0.612***
	(0.058)	(0.079)	(0.079)	(0.158)	(0.162)
N	9299	9299	9299	9053	8762
R^2	0.0361	0.0625	0.0628	0.1131	0.1258

Standard errors in parentheses
"Robust standard errors in parentheses; + z-score"
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.17: Children's cognitive tvip+2012 test z-score: Value added estimation

	(1)
	tvip+2012
Wealth q2	0.090**
	(0.041)
Wealth q3	0.110***
	(0.042)
Wealth q4	0.125***
	(0.047)
Wealth q5	0.154***
	(0.051)
Child's age (mths)	-0.007***
	(0.002)
Male child =1	-0.113***
	(0.026)
Child attends ECC =1	0.198
	(0.121)
Urban HH =1	0.054
	(0.041)
PC's Primary education =1	0.054
	(0.050)
PC's Secondary education =1	0.167***
	(0.051)
PC's Superior education =1	0.198**
	(0.081)
PC's age 20-29 =1	0.030
	(0.040)
PC's age 30-39 =1	0.132***
	(0.046)
PC's age 40+ =1	0.220***
	(0.070)
Younger children =1	-0.019
	(0.030)
Older children =1	-0.052*
	(0.031)
Two-parent family =1	0.038
	(0.029)
PC's LM participation =1	0.042
	(0.027)
tvip+2010	0.318***
	(0.015)
Child's length at birth+	0.017
	(0.016)
Child's weight at birth+	-0.005
	(0.016)
WAIS_Digit Span+	0.000
	(0.002)
WAIS_Vocabulary+	0.007***
	(0.002)
BFI_Extroversion	0.033*
	(0.018)
BFI_Agreeableness	0.003
	(0.023)
BFI_Conscientiousness	-0.001
	(0.025)
BFI_Neuroticism	-0.008
	(0.017)
BFI_Openness	-0.001
	(0.022)
Learning materials score+ 2012	0.046***
	(0.016)
Learning materials score+ 2010	0.033*
	(0.019)
Parent-Child score+ 2012	0.027
	(0.017)
Parent-Child score+ 2010	0.044**
	(0.018)
Responsivity score+ 2012	0.004
	(0.016)
Responsivity score+ 2010	0.011
	(0.019)
Involvement score+2010	0.004
	(0.019)
Constant	-0.385
	(0.250)
N	5164
R^2	0.2195

Standard errors in parentheses
"Robust standard errors in parentheses; + z-score"
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.18: Children's cognitive cbcl1+2012 test z-score: Value added estimation

	(1)
	cbcl1+2012
Wealth q2	-0.061
	(0.051)
Wealth q3	0.019
	(0.052)
Wealth q4	0.002
	(0.053)
Wealth q5	-0.044
	(0.058)
Child's age (mths)	-0.004
	(0.003)
Male child =1	0.016
	(0.029)
Child attends ECC =1	0.092
	(0.091)
Urban HH =1	-0.003
	(0.052)
PC's Primary education =1	-0.099
	(0.074)
PC's Secondary education =1	-0.213***
	(0.074)
PC's Superior education =1	-0.283***
	(0.092)
PC's age 20-29 =1	0.087*
	(0.047)
PC's age 30-39 =1	0.110**
	(0.055)
PC's age 40+ =1	0.155**
	(0.075)
Younger children =1	0.045
	(0.036)
Older children =1	-0.036
	(0.034)
Two-parent family =1	-0.147***
	(0.035)
PC's LM participation =1	-0.000
	(0.031)
cbcl1+2010	0.383***
	(0.018)
Child's length at birth+	-0.013
	(0.019)
Child's weight at birth+	0.018
	(0.019)
WAIS_Digit Span+	-0.002
	(0.002)
WAIS_Vocabulary+	-0.003*
	(0.002)
BFI_Extroversion	-0.015
	(0.021)
BFI_Agreeableness	-0.073***
	(0.028)
BFI_Conscientiousness	-0.028
	(0.028)
BFI_Neuroticism	0.075***
	(0.022)
BFI_Openness	0.059**
	(0.026)
Learning materials score+ 2012	-0.084***
	(0.019)
Learning materials score+ 2010	0.010
	(0.023)
Parent-Child score+ 2012	-0.016
	(0.019)
Parent-Child score+ 2010	-0.035
	(0.022)
Responsivity score+ 2012	-0.126***
	(0.020)
Responsivity score+ 2010	0.088***
	(0.023)
Involvement score+2010	0.024
	(0.022)
Constant	0.596**
	(0.300)
N	3499
R^2	0.2654

Standard errors in parentheses
"Robust standard errors in parentheses; + z-score"
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix 5: Descriptive Statistics for the Estimation of the Production Function for Human Capital

Table A.19: Descriptive statistics age 7-23 mths

Variable	Obs	Mean	Std. Dev.
Child's Cognitive Measures t			
fc_eedp	3805	510.98	135.82
fc_bdi_cog	3804	21.09	5.8
fc_bdi_com	3804	24.71	7.93
fc_bdi_m	3804	55.34	16.17
Child's Non-cognitive Measures t			
fn_bdi_ps	3804	53.06	15.59
fn_bdi_ad	3804	38.58	11.59
fn_asq	2276	33.42	21
Mother's Cognitive Measures			
years_education	3791	11.4	2.97
pc_wais_ds1	3822	5.15	1.3
pc_wais_ds2	3822	3.7	.98
pc_wais_voc	3822	32.91	16.87
Mother's Non-cognitive Measures			
pn_bfi_e	3722	3.56	.73
pn_bfi_a	3722	3.81	.6
pn_bfi_c	3722	3.96	.58
pn_bfi_n	3722	2.94	.83
pn_bfi_o	3722	3.81	.64
pn_psi_pd	3267	31.79	9.63
pn_psi_pcdi	3132	40.81	7.41
Investment Measures			
r_parent_child	3828	5.62	1.4
r_lear_materials	3828	6.19	1.88
r_responsivity	3828	6.35	1.24
r_involvement	3828	3.86	1.7
r_acceptance	3828	3.53	1.19
r_discipline	3828	14.7	3.15
r_program_ccc	3828	.43	.93
Child's Cognitive Measures $t + 1$			
fc_bdi_cog	3719	23.79	6.36
fc_bdi_com	3720	25.97	6.16
fc_bdi_m	3720	26.37	4.78
fc_tadi_c	3737	33.35	6.61
fc_tadi_m	3727	33.61	5
fc_tadi_l	3737	34.24	5.43
fc_tvip_l	3777	16.47	12.24
Child's Non-cognitive Measures $t + 1$			
fn_cbcl_r	3788	20.92	2.9
fn_cbcl_a	3788	11.66	2.99
fn_cbcl_q	3785	17.89	2.62
fn_cbcl_e	3788	13.08	2.35
fn_cbcl_ps	3788	16.16	2.63
fn_cbcl_pa	3788	16.29	1.97
fn_cbcl_ca	3788	24.35	8.33
fn_psi_dc	3118	33.93	9.08
Exclusion restrictions			
wealth	3828	-.08	1.58
unemployment	3828	.08	.03
lwf	3828	12.59	.3
lwm	3828	13.01	.33
p_i_learning	3828	95.51	2.76
Productivity			
male_ch	3828	.51	.50
n_sib	3828	.85	.97
weight_birth	3818	3.33	.55
length_birth	3818	49.32	2.48

Table A.20: Descriptive statistics age 24-47 mths

Variable	Obs	Mean	Std. Dev.
Child's Cognitive Measures t			
fc_tepsi_coo	5995	7.09	3.85
fc_tepsi_l	6019	11.17	7.21
fc_tepsi_m	5995	6.33	2.69
fc_tvip_l	4501	15.39	12.49
Child's Non-cognitive Measures t			
fn_cbcl_r	6054	14.33	2.98
fn_cbcl_a	6054	10.93	2.98
fn_cbcl_q	6054	17.26	2.71
fn_cbcl_e	6054	12.81	2.3
fn_cbcl_ps	6054	10.53	2.72
fn_cbcl_pa	6054	5.94	1.76
fn_cbcl_ca	6054	21.24	7.91
Mother's Cognitive Measures			
years_education	5986	11.31	3
pc_wais_ds1	6060	5.13	1.33
pc_wais_ds2	6060	3.68	1.01
pc_wais_voc	6060	32.88	17.47
Mother's Non-cognitive Measures			
pn_bfi_e	5899	3.51	.74
pn_bfi_a	5899	3.83	.6
pn_bfi_c	5899	4	.58
pn_bfi_n	5899	2.93	.81
pn_bfi_o	5899	3.78	.64
pn_psi_pd	5239	31.89	9.67
pn_psi_pcdi	5050	40.12	8.09
Investment Measures			
r_parent_child	6063	5.93	1.28
r_lear_materials	6063	6.54	1.77
r_responsivity	6063	6.02	1.45
r_involvement	6063	3.44	1.83
r_acceptance	6063	3.46	1.21
r_discipline	6063	14.95	3.11
r_program_ccc	6063	.15	.54
Child's Cognitive Measures $t+1$			
fc_bdi_cog	5936	33.33	5.51
fc_bdi_com	5938	33.02	5.07
fc_bdi_m	5937	34.72	4.7
fc_tadi_c	5961	45.55	5.05
fc_tadi_m	5964	42.72	3.84
fc_tadi_l	5972	46.92	5.83
fc_tvip_l	6003	39.53	17.69
Child's Non-cognitive Measures $t+1$			
fn_cbcl_r	6020	20.76	3.1
fn_cbcl_a	6020	11.89	3.1
fn_cbcl_q	6016	17.68	2.75
fn_cbcl_e	6020	13.25	2.39
fn_cbcl_ps	6020	16.29	2.57
fn_cbcl_pa	6020	16.37	2.25
fn_cbcl_ca	6020	26.64	8.29
fn_psi_dc	5025	33.9	9.26
Exclusion restrictions			
wealth	6063	-.07	1.55
unemployment	6063	.08	.03
lwf	6063	12.58	.3
lwm	6063	13	.33
p_i_learning	6063	95.53	2.74
Productivity			
male_ch	6063	.50	.50
n_sib	6063	.98	1
weight_birt	6028	3.34	.53
length_birth	6027	49.46	2.45

Table A.21: Descriptive statistics Multiple Investments age 24-47 mths

Variable	Obs	Mean	Std. Dev.
Vocalises	6056	.95	.21
Speech	6056	.96	.2
Verbal	6056	.9	.3
Expresses	6056	.9	.3
Praises	6056	.75	.43
Caresses	6057	.73	.44
PraisetoChild	6057	.84	.37
Shout	6056	.87	.34
Annoyance	6057	.85	.36
Slaps	6057	.89	.31
Scold	6056	.85	.36
Visual	6057	.82	.39
Talks	6056	.79	.41
Encourages	6056	.69	.46
Toys	6057	.44	.5
Period	6057	.33	.47
Challenge	6057	.38	.49
SpecialPlace	6048	.89	.31
MuscularAct	6042	.81	.39
PushPull	6053	.87	.34
WheelsToys	6051	.87	.33
RolePlay	6051	.91	.29
LearEquip	6044	.69	.46
Musical	6041	.79	.41
Books3	6024	.74	.44
LookBooks	6063	.77	.42
Stories	6063	.79	.4
Sing	6063	.96	.19
GoOut	6063	.81	.39
Play	6063	.99	.1
Drawing	6063	.97	.17

Table A.22: Descriptive statistics age 48-58 mths

Variable	Obs	Mean	Std. Dev.
Child's Cognitive Measures t			
fc_tepsi_coo	1300	12.61	2.63
fc_tepsi_l	1303	19.9	4.09
fc_tepsi_m	1300	9.44	2.12
fc_tvip_l	1302	29.86	13.99
Child's Non-cognitive Measures t			
fn_cbcl_r	1306	14.28	3.07
fn_cbcl_a	1306	11	3.08
fn_cbcl_q	1306	17.14	2.64
fn_cbcl_e	1306	12.9	2.38
fn_cbcl_ps	1306	10.8	2.58
fn_cbcl_pa	1306	6.01	1.82
fn_cbcl_ca	1306	22.11	8.05
Mother's Cognitive Measures			
esc	1287	11.16	3.13
pc_wais_ds1	1305	5.1	1.28
pc_wais_ds2	1305	3.65	1
pc_wais_voc	1305	33.34	17.38
Mother's Non-cognitive Measures			
pn_bfi_e	1273	3.5	.73
pn_bfi_a	1273	3.88	.6
pn_bfi_c	1273	4.01	.57
pn_bfi_n	1273	2.99	.84
pn_bfi_o	1273	3.81	.64
pn_psi_pd	1108	31.89	9.86
pn_psi_pcdi	1082	39.13	8.64
Investment Measures			
r_parent_child	1307	6.11	1.17
r_lear_materials	1307	6.66	1.64
r_responsivity	1307	6.03	1.44
r_involvement	1307	3.43	1.98
r_acceptance	1307	3.4	1.27
r_discipline	1307	15.08	3.17
r_program_ccc	1307	.09	.42
Child's Cognitive Measures $t+1$			
fc_bdi_cog	1287	37.27	3.18
fc_bdi_com	1287	36.45	3.7
fc_bdi_m	1287	38.31	2.48
fc_tadi_c	1282	50.04	2.67
fc_tadi_m	1276	45.78	1.97
fc_tadi_l	1291	52.73	2.98
fc_tvip_l	1298	55.82	15.95
Child's Non-cognitive Measures $t+1$			
fn_cbcl_r	1295	18.33	4.11
fn_cbcl_a	1295	13.78	2.44
fn_cbcl_q	1295	16.66	3.21
fn_cbcl_e	1295	11.68	3.2
fn_cbcl_ps	1295	16.59	2.8
fn_cbcl_pa	1295	14.45	3.94
fn_cbcl_ca	1295	35.42	3.02
fn_psi_dc	1082	33.72	9.7
Exclusion restrictions			
wealth	1307	-.05	1.55
unemployment	1307	.08	.03
lwf	1307	12.6	.32
lwm	1307	13.02	.36
p_i_learning	1307	95.57	2.64
Productivity			
male_ch	1307	.5	.5
n_sib	1307	1.07	.99
weight_birth	1304	3.36	.52
length_birth	1304	49.48	2.61

Table A.23: Total Variance in Measures: Signal and Noise

Measurements	%Signal	%Noise
Child's Cognitive Measures t		
fc_tepsi_coo	0.884	0.116
fc_tepsi_l	0.908	0.092
fc_tepsi_m	0.793	0.207
fc_tvip_l	0.756	0.244
Child's Non-cognitive Measures t		
fn_cbcl_r	0.924	0.076
fn_cbcl_a	0.826	0.174
fn_cbcl_q	0.658	0.342
fn_cbcl_e	0.683	0.317
fn_cbcl_ps	0.679	0.321
fn_cbcl_pa	0.635	0.365
fn_cbcl_ca	0.791	0.209
Mother's Cognitive Measures		
years_education	0.823	0.177
pc_wais_ds1	0.711	0.289
pc_wais_ds2	0.821	0.179
pc_wais_voc	0.803	0.197
Mother's Non-cognitive Measures		
pn_bfi_e	0.803	0.197
pn_bfi_a	0.528	0.472
pn_bfi_c	0.962	0.038
pn_bfi_n	0.981	0.019
pn_bfi_o	0.914	0.086
pn_psi_pd	0.853	0.147
pn_psi_pcdi	0.758	0.242
Investment Measures		
Total		
r_parent_child	0.794	0.206
r_lear_materials	0.911	0.089
r_responsivity	0.493	0.507
r_involvement	0.473	0.527
r_acceptance	0.076	0.924
r_discipline	0.001	0.999
r_program_ccc	0.027	0.973
Monetary and Time		
SpecialPlace	0.822	0.178
WheelsToys	0.845	0.155
LearEquip	0.842	0.158
Books3	0.738	0.262
LookBooks	0.854	0.146
Stories	0.856	0.144
GoOut	0.212	0.788
Drawing	0.353	0.647
Cognitive and Emotional		
RolePlay	0.822	0.178
PushPull	0.786	0.214
LearEquip	0.848	0.152
Musical	0.808	0.192
Vocalises	0.854	0.146
Visual	0.756	0.244
Praises	0.619	0.381
Caresses	0.313	0.687
Child's Cognitive Measures $t+1$		
fc_bdi_cog	0.896	0.104
fc_bdi_com	0.850	0.150
fc_bdi_m	0.816	0.184
fc_tadi_c	0.860	0.140
fc_tadi_m	0.613	0.387
fc_tadi_l	0.881	0.119
fc_ppvt_l	0.744	0.256
Child's Non-cognitive Measures $t+1$		
fn_cbcl_r	0.894	0.106
fn_cbcl_a	0.834	0.166
fn_cbcl_q	0.727	0.273
fn_cbcl_e	0.766	0.234
fn_cbcl_ps	0.708	0.292
fn_cbcl_pa	0.683	0.317
fn_cbcl_ca	0.883	0.117
fn_psi_dc	0.440	0.560

Table A.24: Factor loadings from the measurement error system

	Loadings									
						Exclusion Restrictions				
Measures	Child's Cognitive	Child's Non-cognitive	Maternal Cognitive	Maternal Non-cognitive	Unobserved Heterogeneity in Investment Policy Equation	Wealth	Unemployment	Wage Female	Wage Male	Price Index Materials and Activities
Child's Cognitive										
fc_tepsi_coo	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
fc_tepsi_l	1.046	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
fc_tepsi_m	0.796	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
fc_tvip_l	0.672	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Child's Non-cognitive										
fn_cbcl_r	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
fn_cbcl_a	0.000	0.785	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
fn_cbcl_q	0.000	0.617	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
fn_cbcl_e	0.000	0.609	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
fn_cbcl_ps	0.000	0.641	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
fn_cbcl_pa	0.000	0.613	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
fn_cbcl_ca	0.000	0.738	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maternal Cognitive										
years_education	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pc_wais_ds1	0.000	0.000	0.629	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pc_wais_ds2	0.000	0.000	0.807	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pc_wais_voc	0.000	0.000	0.844	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maternal Non-cognitive										
pn_bfi_e	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
pn_bfi_a	0.000	0.000	0.000	0.487	0.000	0.000	0.000	0.000	0.000	0.000
pn_bfi_c	0.000	0.000	0.000	2.357	0.000	0.000	0.000	0.000	0.000	0.000
pn_bfi_n	0.000	0.000	0.000	3.000	0.000	0.000	0.000	0.000	0.000	0.000
pn_bfi_o	0.000	0.000	0.000	1.479	0.000	0.000	0.000	0.000	0.000	0.000
pn_psi_pd	0.000	0.000	0.000	1.074	0.000	0.000	0.000	0.000	0.000	0.000
pn_psi_pcdi	0.000	0.000	0.000	0.816	0.000	0.000	0.000	0.000	0.000	0.000
Investment										
r_parent_child	0.781	0.073	0.055	0.282	1.000	0.095	-0.041	-0.119	0.124	0.055
r_lear_materials	1.189	0.111	0.084	0.429	1.521	0.144	-0.063	-0.181	0.189	0.083
r_responsivity	0.395	0.037	0.028	0.143	0.506	0.048	-0.021	-0.060	0.063	0.028
r_involvement	0.379	0.035	0.027	0.137	0.484	0.046	-0.020	-0.058	0.060	0.027
r_acceptance	0.117	0.011	0.008	0.042	0.150	0.014	-0.006	-0.018	0.019	0.008
r_discipline	0.016	0.001	0.001	0.006	0.020	0.002	-0.001	-0.002	0.002	0.001
r_program_ccc	0.068	0.006	0.005	0.025	0.087	0.008	-0.004	-0.010	0.011	0.005

Figure A.0.2: Kernel densities of latent traits: One investment input, 7-23mths, Specification b

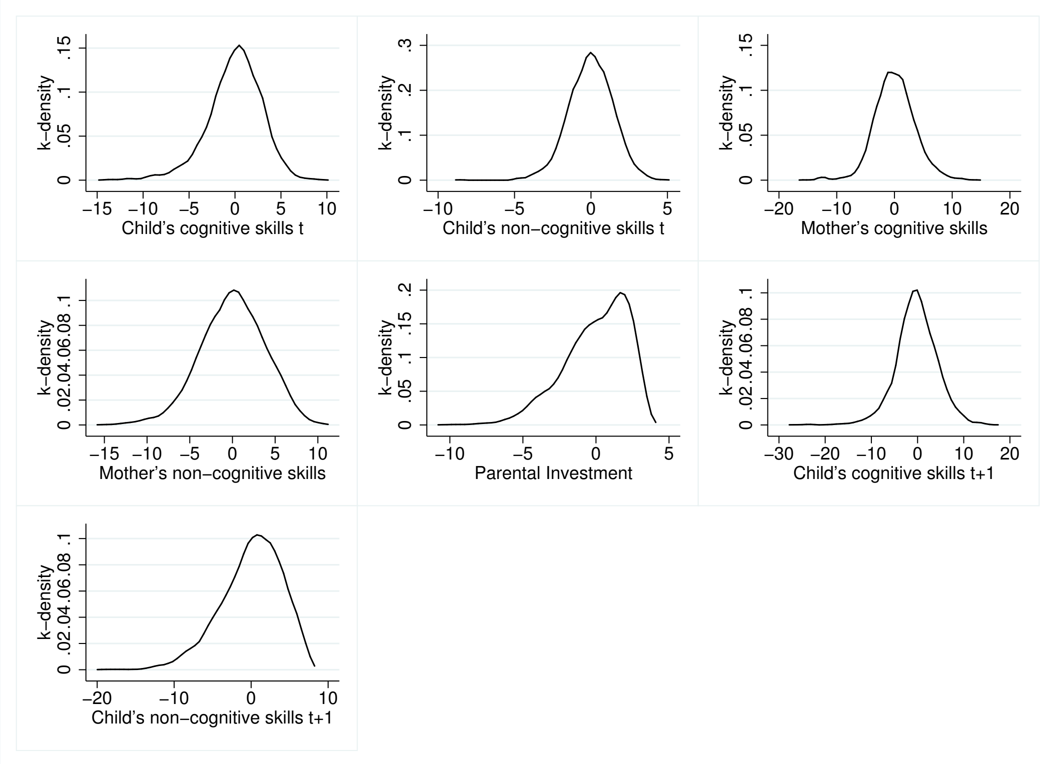
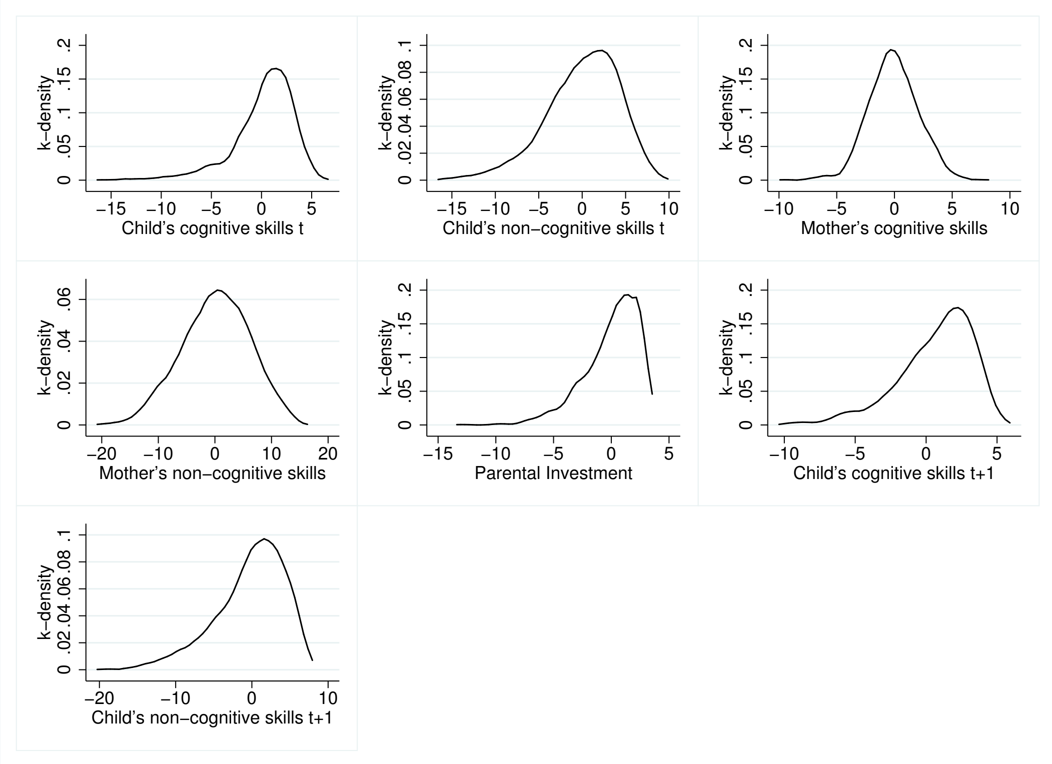


Figure A.0.3: Kernel densities of latent traits: One investment input, 48-58mths, Specification b



Appendix 6: Maternal Investment and Beliefs

Figure A.0.4: Maternal Investment: time and didactic materials

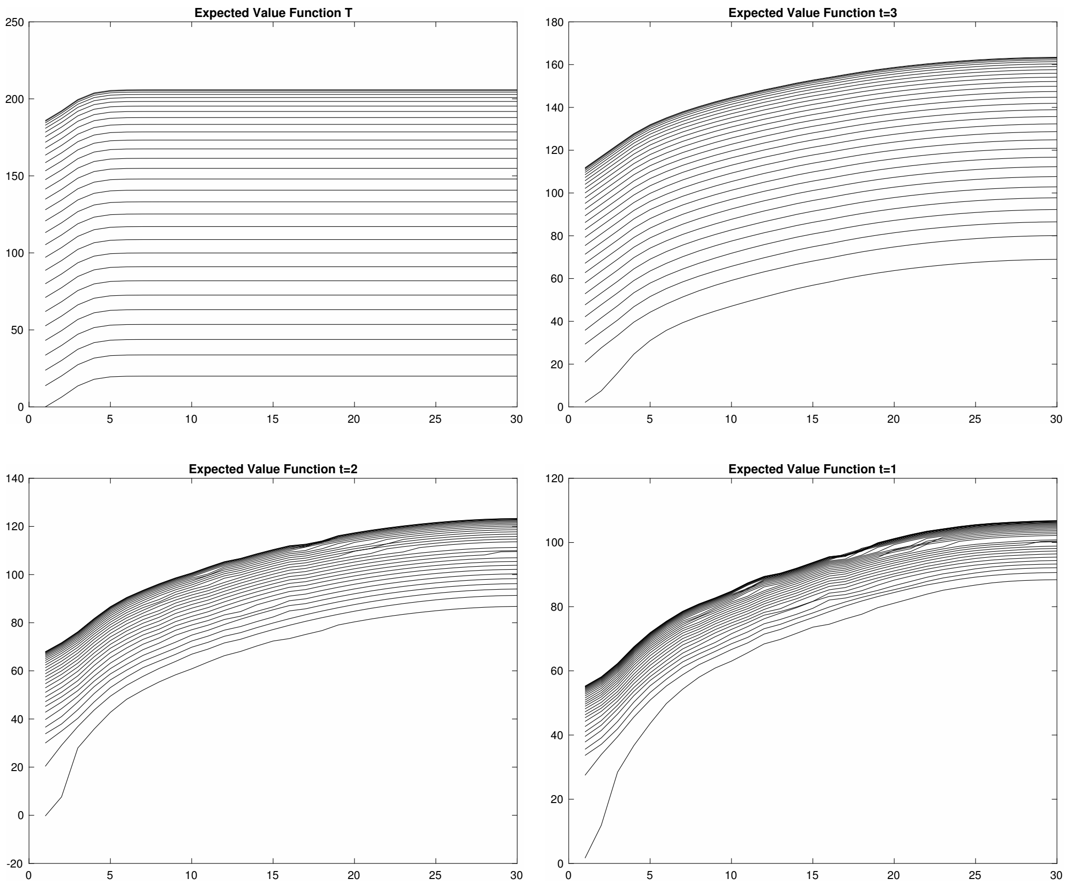


Figure A.0.5: Expected Value Functions for periods T to $t = 1$

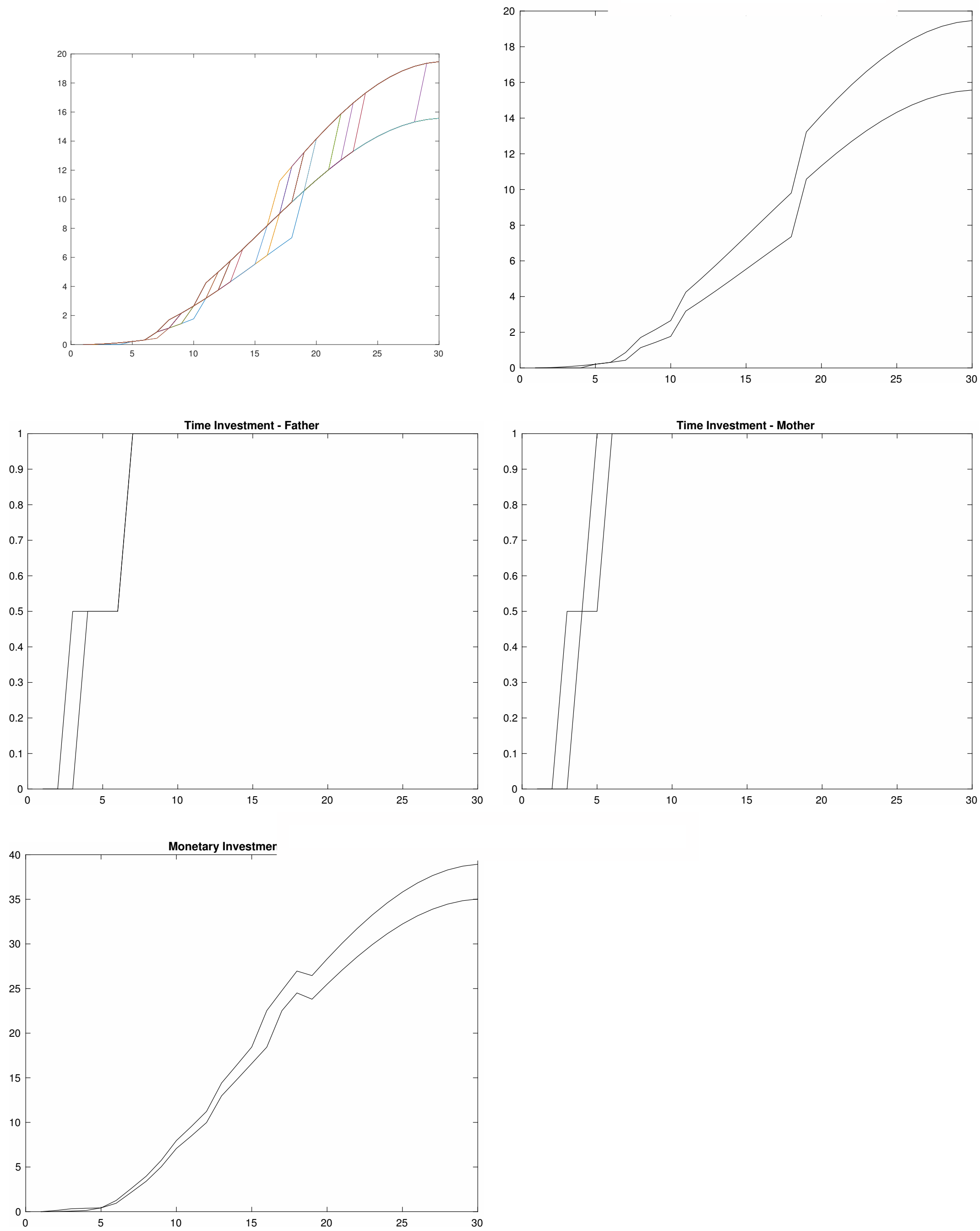


Figure A.0.6: Solution for the optimal family decisions, x axis is the child's cognitive skills